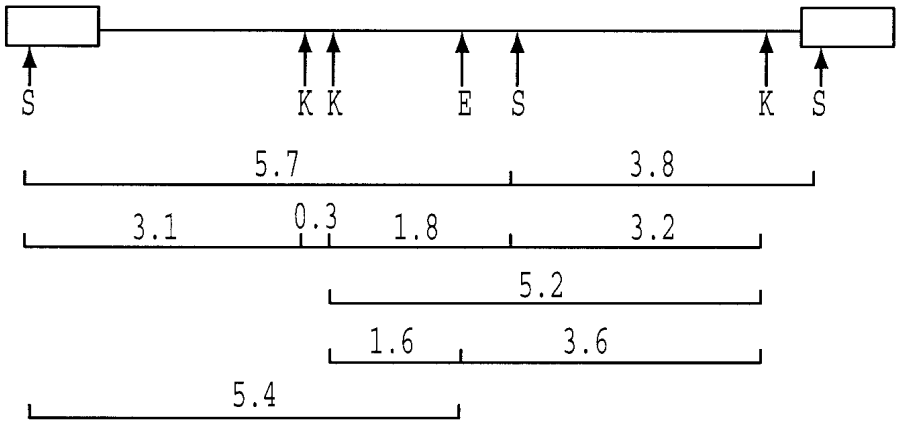


FIG. 1

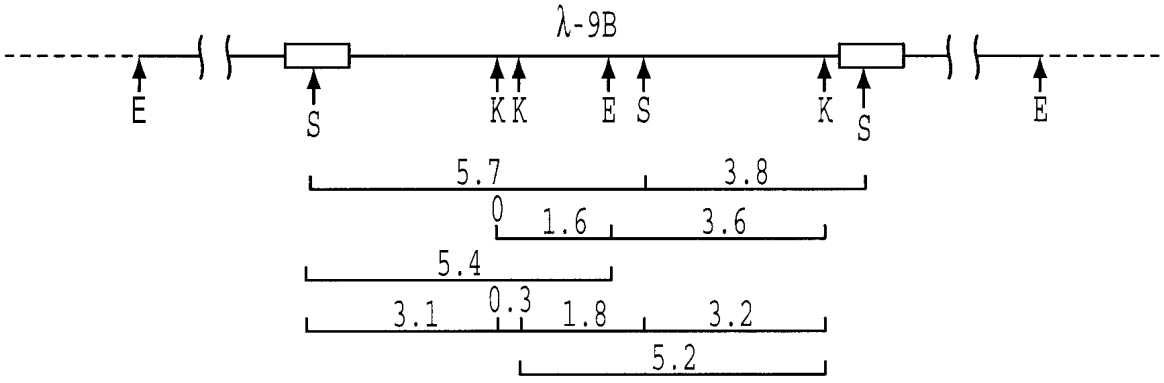


FIG. 2

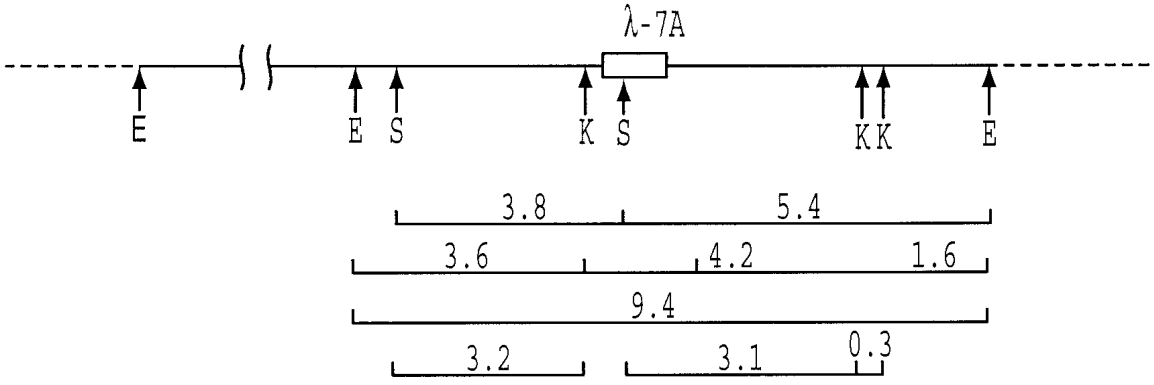


FIG. 3

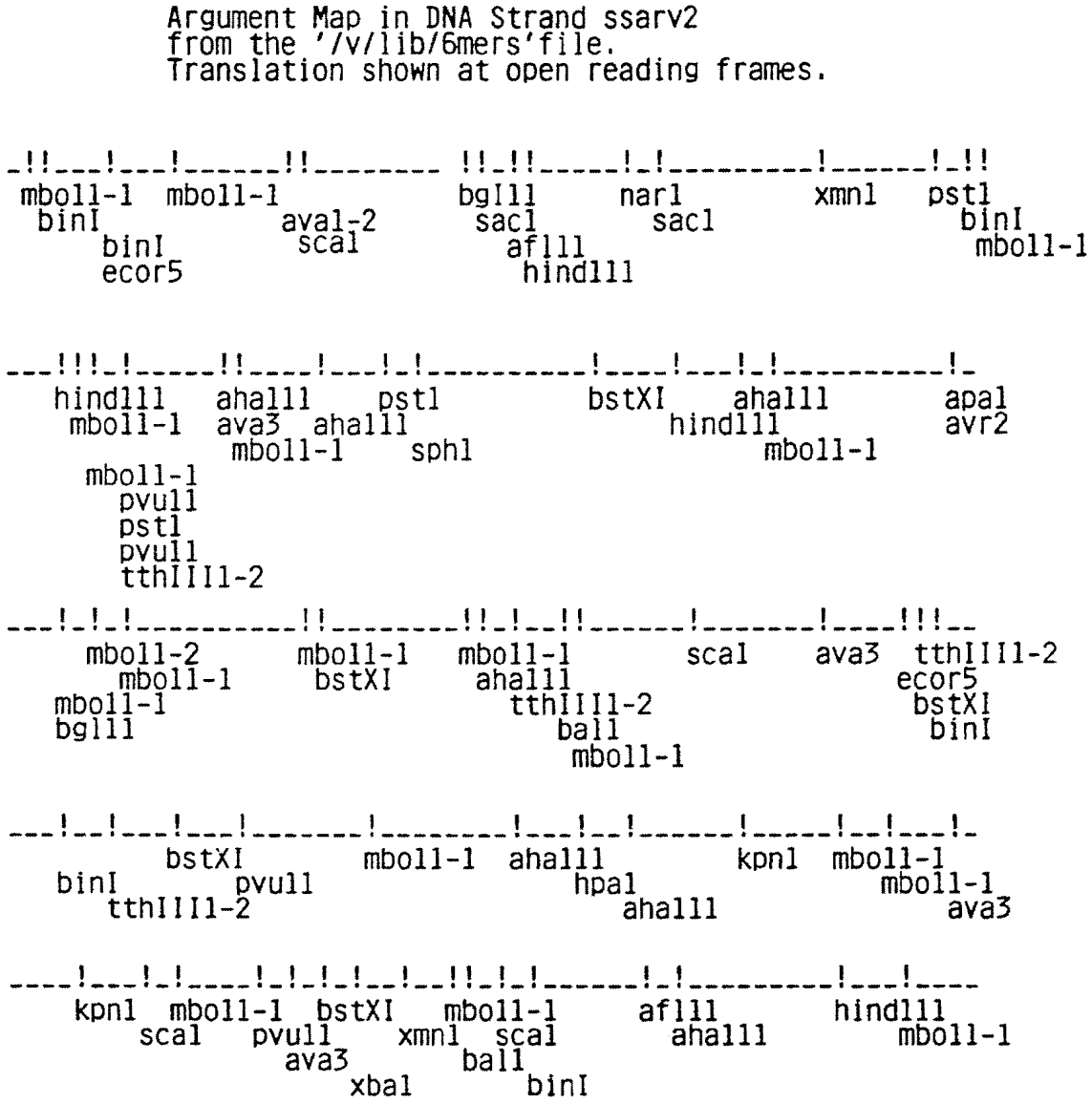

Argument Map in DNA Strand ssarv2
from the '/v/lib/6mers'file.
Translation shown at open reading frames.
mboll-1
binI
mboll-1
aval-2
scal
binI
ecor5
bgl11
sac1
afl11
hindlll
nar1
sac1
xmn1
pst1
binI
mboll-1
hindlll
mboll-1
pvull
pstl
pvull
tthIII-2
ahalll
ava3
mboll-1
ahalll
pstl
sphl
bstXI
hindlll
ahalll
mboll-1
apal
avr2
mboll-2
mboll-1
bgl11
mboll-1
bstXI
mboll-1
ahalll
tthIII-2
ball
mboll-1
scal
ecor5
bstXI
binI
ava3
tthIII-2
bstXI
pvull
tthIII-2
mboll-1
ahalll
hpal
ahalll
kpnl
mboll-1
mboll-1
ava3
kpnl
scal
mboll-1
pvull
ava3
xbal
bstXI
xmn1
mboll-1
ball
binI
scal
afl11
ahalll
hindlll
mboll-1

FIG. 4A

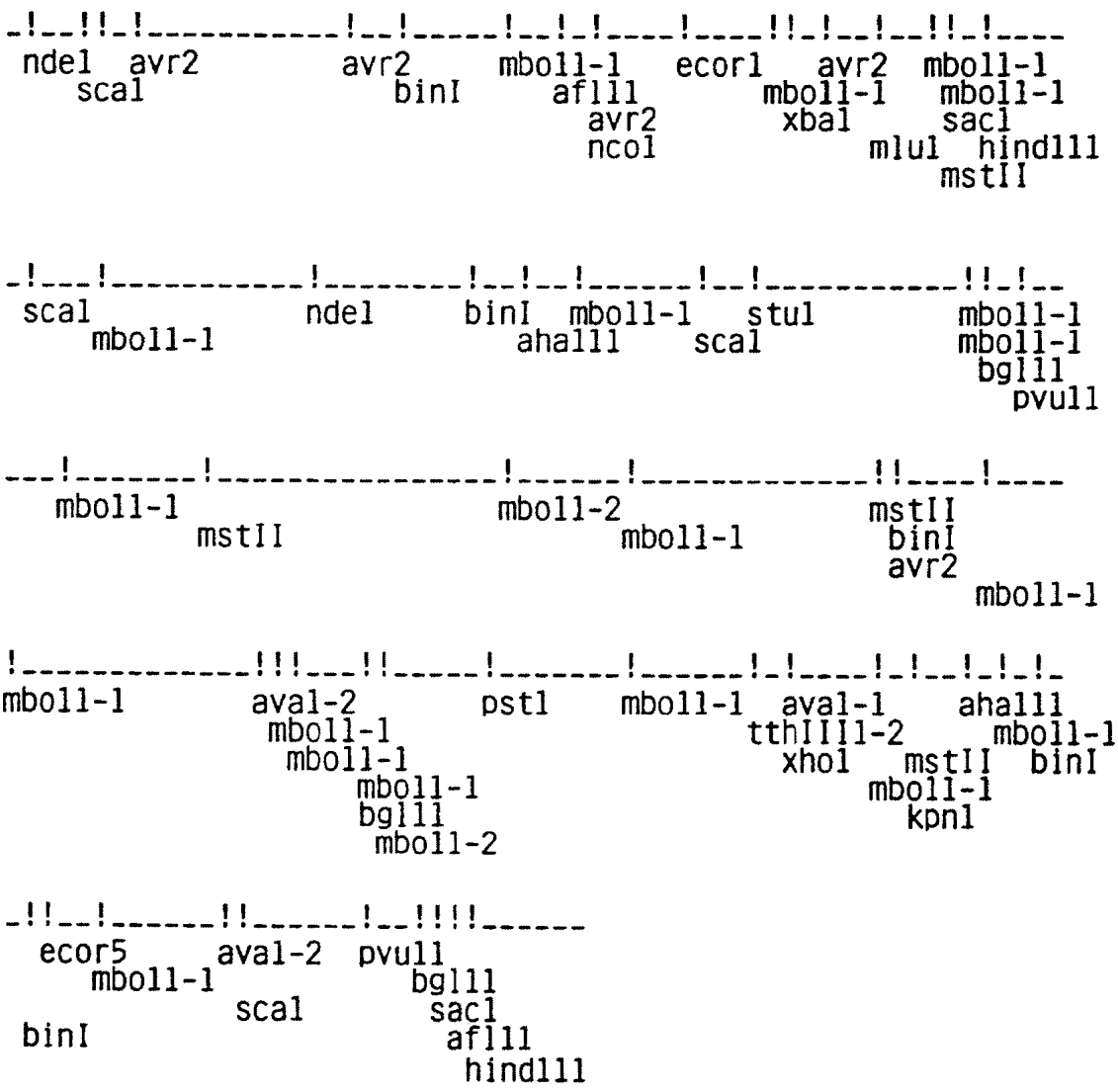

ndeI avr2
scaI
avr2
binI
mboII-1
aflII
avr2
ncoI
ecorI
mboII-1
xbaI
avr2
sacI
mluI
mstII
mboII-1
hindIII
scaI
mboII-1
ndeI
binI
ahaIII
mboII-1
scaI
stuI
mboII-1
mboII-1
bglII
pvuII
mboII-1
mstII
mboII-2
mboII-1
mstII
binI
avr2
mboII-1
mboII-1
aval-2
mboII-1
mboII-1
mboII-1
bglII
mboII-2
pstI
mboII-1
aval-1
tthIII-2
xhoI
mstII
mboII-1
kpnI
ahaIII
mboII-1
binI
ecor5
mboII-1
scaI
binI
aval-2
pvuII
bglII
sacI
aflII
hindIII

FIG. 4B

```
  1 CTGGAAGGGCTAATTTGGTCCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCACAC
    GACCTTCCCGATTAAACCAGGGTTTCTTCTGTTCTCTAGGAACTAGACACCTAGATGGTGTG

    26 mboll, 50 binl,

 63 ACAAGGCTACTTCCCTGATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCACT
    TGTTCCGATGAAGGGACTAACCGTCTTAATGTGTGGTCCCGGTCCCTAGTCTATAGGTGA

    107 binl, 113 ecor5,

123 GACCTTTGGATGGTGCTTCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCCAA
    CTGGAAACCTACCACGAAGTTCGATCATGGTCAACTCGGTCTCTTCCATCTTCTCCGGTT

    172 mboll,

183 TGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCGGA
    ACTTCCTCTCTTGTTGTCGAACAATGTGGGATACTCGGACGTACCCTACCTCCTGCGCCT

243 GAAAGAAGTGTTAGTGTGGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGAGA
    CTTTCTTCACAATCACACCTCCAAACTGTCGTTTGATCGTAAAGTAGTGTACCGGGCTCT

    296 aval,

303 GCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGC
    CGACGTAGGCCTCATGATGTTTCTGACGACTGTAGCTCGAAAGATGTTCCCTGAAAGGCG

    314 scal,

363 TGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATGC
    ACCCCTGAAAGGTCCCTCCGCACCGGACCCGCCCTGACCCCTCACCGCAGGGAGTCTACG

423 TGCATATAAGCAGACTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAG
    ACGTATATTCGTCTGACGAAAAACGGACATGACCCAGAGAGACCAATCTGGTCTAGACTC

    474 bglll,

483 CCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTT
    GGACCCTCGAGAGACCGATTGATCCCTTGGGTGACGAATTCGGAGTTATTTCGAACGGAA

    488 sacl, 518 aflll, 532 hindlll,

543 GAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCA
    CTCACGAAGTTCATCACACACGGGCAGACAACACACTGAGACCATTGATCTCTAGGGAGT

603 GACCCTTTTAGTCAGTGTGGAAAAATCTCTAGCAGTGGCGCCCGAACAGGGACGCGAAAG
    CTGGGAAAATCAGTCACACCTTTTTAGAGATCGTCACCGCGGGCTTGTCCCTGCGCTTTC

    639 narl,

663 CGAAAGTAGAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACAG
    GCTTTCATCTTGGTCTCCTCGAGAGAGCTGCGTCCTGAGCCGAACGACTTCGCGCGTGTC

    680 sacl,

723 CAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAATTTTTGACTAGCGGAGGCTAGAAG
    GTTCTCCGCTCCCCGCCGCTGACCACTCATGCGGTTAAAAACTGATCGCCTCCGATCTTC

                MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspLysTrpGlu GAG
783 GAGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAATGGGAA
    CTCTCTCTCTACCCACGCTCTCGCAGCCATAATTCGCCCCCTCTTAATCTATTTACCCTT
```

FIG. 4C

```
                LysIleArgLeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAla
843             AAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAGTTAAAACATATAGTATGGGCA
                TTTTAAGCCAATTCCGGTCCCCCTTTCTTTTTTATATTCAATTTTGTATATCATACCCGT

                SerArgGluLeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCys
903             AGCAGGGAGCTAGAACGATTCGCAGTCAATCCTGGCCTGTTAGAAACATCAGAAGGCTGC
                TCGTCCCTCGATCTTGCTAAGCGTCAGTTAGGACCGGACAATCTTTGTAGTCTTCCGACG

                959 pstI,

                ArgGlnIleLeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSer
963             AGACAAATATTGGGACAGCTACAGCCATCCCTTCAGACAGGATCAGAAGAACTTAGATCA
                TCTGTTTATAACCCTGTCGATGTCGGTAGGGAAGTCTGTCCTAGTCTTCTTGAATCTAGT

                1002 binI, 1008 mboII,

                LeuTyrAsnThrValAlaThrLeuTyrCysValHisGlnArgIleAspValLysAspThr
1023            TTATATAATACAGTAGCAACCCTCTATTGTGTACATCAAAGGATAGATGTAAAAGACACC
                AATATATTATGTCATCGTTGGGAGATAACACATGTAGTTTCCTATCTACATTTTCTGTGG

                LysGluAlaLeuGluLysIleGluGluGluGlnAsnLysSerLysLysLysAlaGlnGln
1083            AAGGAAGCTTTAGAGAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAGGCACAGCAA
                TTCCTTCGAAATCTCTTCTATCTCCTTCTCGTTTTGTTTTCATTCTTTTTCCGTGTCGTT

                1087 hindIII, 1097 mboII, 1107 mboII,

                                                                    p25
                AlaAlaAlaAlaAlaGlyThrGlyAsnSerSerGlnValSerGlnAsnTyrProIleVal
1143            GCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTCAGCCAAAATTACCCTATAGTG
                CGTCGTCGACGTCGACCGTGTCCTTTGTCGTCGGTCCAGTCGGTTTTAATGGGATATCAC

                1147 pvuII, 1150 pstI, 1153 pvuII, 1156 tthIIII,

                GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrp
1203            CAGAACCTACAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG
                GTCTTGGATGTCCCCGTTTACCATGTAGTCCGGTATAGTGGATCTTGAAATTTACGTACC

                1250 ahaIII, 1255 ava3,

                ValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu
1263            GTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA
                CATTTTCATCATCTTCTTTTCCGAAAGTCGGGTCTTCATTATGGGTACAAAAGTCGTAAT

                1275 mboII,

                SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGln
1323            TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAA
                AGTCTTCCTCGGTGGGGTGTTCTAAATTTGTGGTACGATTTGTGTCACCCCCCTGTAGTT

                1346 ahaIII,

                AlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal
1383            GCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG
                CGTCGGTACGTTTACAATTTTCTCTGATAGTTACTCCTTCGACGTCTTACCCTATCTCAC

                1423 pstI,

                HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAsp
1443            CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGAC
                GTAGGTCACGTACGTCCCGGATAACGTGGTCCGGTTTACTCTCTTGGTTCCCCTTCACTG

                1451 sphI,
```

FIG. 4D

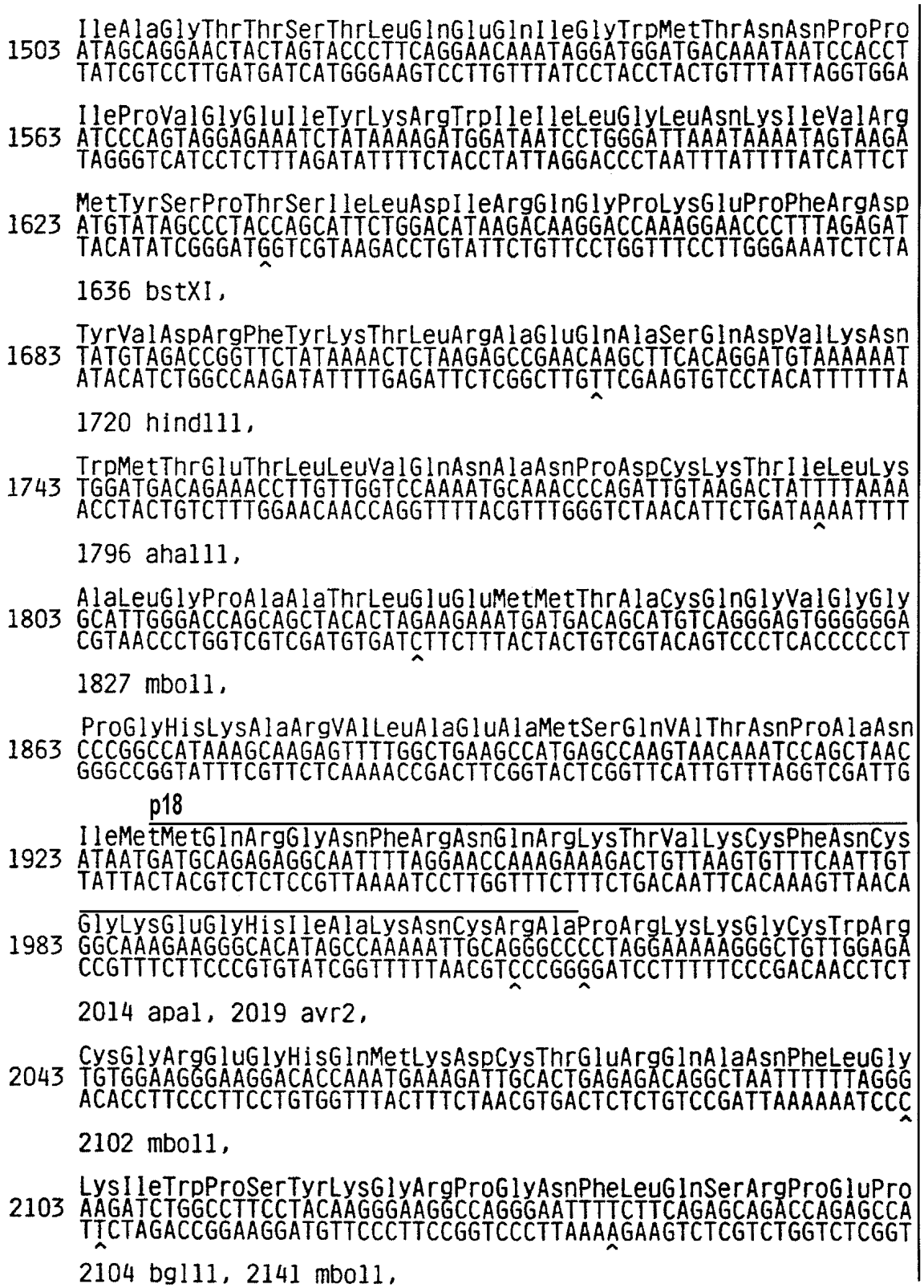

IleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProPro
1503 ATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT
TATCGTCCTTGATGATCATGGGAAGTCCTTGTTTATCCTACCTACTGTTTATTAGGTGGA
IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArg
1563 ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGA
TAGGGTCATCCTCTTTAGATATTTTCTACCTATTAGGACCCTAATTTATTTTATCATTCT
MetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp
1623 ATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT
TACATATCGGGATGGTCGTAAGACCTGTATTCTGTTCCTGGTTTCCTTGGGAAATCTCTA
1636 bstXI,
TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsn
1683 TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAAT
ATACATCTGGCCAAGATATTTTGAGATTCTCGGCTTGTTCGAAGTGTCCTACATTTTTTA
1720 hindIII,
TrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys
1743 TGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA
ACCTACTGTCTTTGGAACAACCAGGTTTTACGTTTGGGTCTAACATTCTGATAAAATTTT
1796 ahaIII,
AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGly
1803 GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGA
CGTAACCCTGGTCGTCGATGTGATCTTCTTTACTACTGTCGTACAGTCCCTCACCCCCCT
1827 mboII,
ProGlyHisLysAlaArgVAlLeuAlaGluAlaMetSerGlnVAlThrAsnProAlaAsn
1863 CCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCCATGAGCCAAGTAACAAATCCAGCTAAC
GGGCCGGTATTTCGTTCTCAAAACCGACTTCGGTACTCGGTTCATTGTTTAGGTCGATTG
p18
IleMetMetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCys
1923 ATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGACTGTTAAGTGTTTCAATTGT
TATTACTACGTCTCTCCGTTAAAATCCTTGGTTTCTTTCTGACAATTCACAAAGTTAACA
GlyLysGluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpArg
1983 GGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAGA
CCGTTTCTTCCCGTGTATCGGTTTTTAACGTCCCGGGGATCCTTTTTCCCGACAACCTCT
2014 apaI, 2019 avrII,
CysGlyArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGly
2043 TGTGGAAGGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGG
ACACCTTCCCTTCCTGTGGTTTACTTTCTAACGTGACTCTCTGTCCGATTAAAAAATCCC
2102 mboII,
LysIleTrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluPro
2103 AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA
TTCTAGACCGGAAGGATGTTCCCTTCCGGTCCCTTAAAAGAAGTCTCGTCTGGTCTCGGT
2104 bglII, 2141 mboII,

FIG. 4E

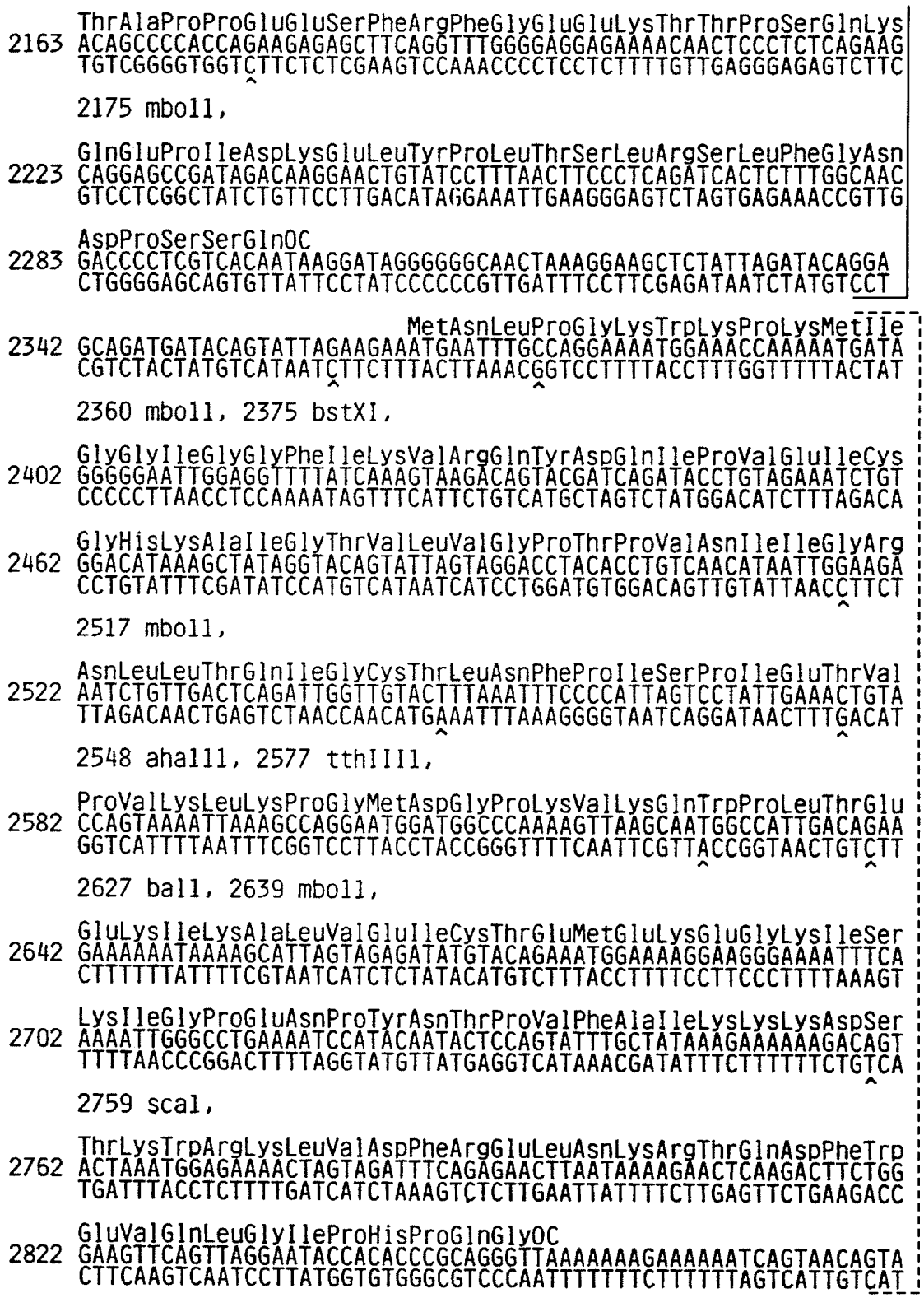

ThrAlaProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLys
2163 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG
TGTCGGGGTGGTCTTCTCTCGAAGTCCAAACCCCTCCTCTTTTGTTGAGGGAGAGTCTTC
2175 mboII,
GlnGluProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsn
2223 CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC
GTCCTCGGCTATCTGTTCCTTGACATAGGAAATTGAAGGGAGTCTAGTGAGAAACCGTTG
AspProSerSerGlnOC
2283 GACCCCTCGTCACAATAAGGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGA
CTGGGGAGCAGTGTTATTCCTATCCCCCCGTTGATTTCCTTCGAGATAATCTATGTCCT
MetAsnLeuProGlyLysTrpLysProLysMetIle
2342 GCAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATA
CGTCTACTATGTCATAATCTTCTTTACTTAAACGGTCCTTTTACCTTTGGTTTTTACTAT
2360 mboII, 2375 bstXI,
GlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProValGluIleCys
2402 GGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTACGATCAGATACCTGTAGAAATCTGT
CCCCCTTAACCTCCAAAATAGTTTCATTCTGTCATGCTAGTCTATGGACATCTTTAGACA
GlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArg
2462 GGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGA
CCTGTATTTCGATATCCATGTCATAATCATCCTGGATGTGGACAGTTGTATTAACCTTCT
2517 mboII,
AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal
2522 AATCTGTTGACTCAGATTGGTTGTACTTTAAATTTCCCCATTAGTCCTATTGAAACTGTA
TTAGACAACTGAGTCTAACCAACATGAAATTTAAAGGGGTAATCAGGATAACTTTGACAT
2548 ahaIII, 2577 tthIII,
ProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu
2582 CCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAGCAATGGCCATTGACAGAA
GGTCATTTTAATTTCGGTCCTTACCTACCGGGTTTTCAATTCGTTACCGGTAACTGTCTT
2627 ball, 2639 mboII,
GluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSer
2642 GAAAAAATAAAAGCATTAGTAGAGATATGTACAGAAATGGAAAAGGAAGGGAAAATTTCA
CTTTTTTATTTTCGTAATCATCTCTATACATGTCTTTACCTTTTCCTTCCCTTTTAAAGT
LysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer
2702 AAAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCTATAAAGAAAAAAGACAGT
TTTTAACCCGGACTTTTAGGTATGTTATGAGGTCATAAACGATATTTCTTTTTTCTGTCA
2759 scaI,
ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp
2762 ACTAAATGGAGAAAACTAGTAGATTTCAGAGAACTTAATAAAAGAACTCAAGACTTCTGG
TGATTTACCTCTTTTGATCATCTAAAGTCTCTTGAATTATTTTCTTGAGTTCTGAAGACC
GluValGlnLeuGlyIleProHisProGlnGlyOC
2822 GAAGTTCAGTTAGGAATACCACACCCGCAGGGTTAAAAAAAGAAAAAATCAGTAACAGTA
CTTCAAGTCAATCCTTATGGTGTGGGCGTCCCAATTTTTTTCTTTTTTAGTCATTGTCAT

FIG. 4F

```
2882 TTGGATGTGGGTGATGCATACTTTTCAGTTCCCTTAGATAAAGACTTTAGAAAGTATACTG
     AACCTACACCCACTACGTATGAAAAGTCAAGGGAATCTATTTCTGAAATCTTTCATATGAC

     2895 ava3,

                                        MetArgHisGlnGlyLeuAspIleSerThrMetTrp  POL
2943 CATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGG
     GTAAATGGTATGGATCATATTTGTTACTCTGTGGTCCCTAATCTATAGTCATGTTACACC

     2985 ecor5,

     LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeu
3003 CTGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTA
     GACGGTGTCCCTACCTTTCCTAGTGGTCGTTATAAGGTTTCATCGTACTGTTTTTAGAAT

     3003 tthIII1, 3006 bstXI, 3021 binI,

     GluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyr
3063 GAGCCTTTTAGAAAACAGAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTAT
     CTCGGAAAATCTTTTGTCTTAGGTCTGTATCAATAGATAGTTATGTACCTACTAAACATA

     ValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHis
3123 GTAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAACTGAGACAGCAT
     CATCCTAGACTGAATCTTTATCCCGTCGTATCTTGTTTTTATCTCCTTGACTCTGTCGTA

     3126 binI, 3171 tthIII1,

     LeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu
3183 CTGTTGAGGTGGGGATTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTT
     GACAACTCCACCCCTAAATGGTGTGGTCTGTTTTTTGTAGTCTTTCTTGGAGGTAAGGAA

     3234 bstXI,

     TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMetLeuProGlu
3243 TGGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAATGCTGCCAGAA
     ACCTACCCAATACTTGAGGTAGGACTATTTACCTGTCATGTCGGATATTACGACGGTCTT

     LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer
3303 AAAGACAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGAAAATTGAATTGGGCAAGT
     TTTCTGTCGACCTGACAGTTACTGTATGTCTTCAATCACCCTTTTAACTTAACCCGTTCA

     3308 pvuII,

     GlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThrLysAla
3363 CAGATTTATGCAGGGATTAAAGTAAAGCAGTTATGTAAACTCCTTAGAGGAACCAAAGCA
     GTCTAAATACGTCCCTAATTTCATTTCGTCAATACATTTGAGGAATCTCCTTGGTTTCGT

     LeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu
3423 CTAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGGGAG
     GATTGTCTTCATTATGGTGATTGTCTTCTTCGTCTCGATCTTGACCGTCTTTTGTCCCTC

     3447 mboII,

     IleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeuValAlaGlu
3483 ATTCTAAAAGAACCAGTACATGAAGTATATTATGACCCATCAAAAGACTTAGTAGCAGAA
     TAAGATTTTCTTGGTCATGTACTTCATATAATACTGGGTAGTTTTCTGAATCATCGTCTT

     IleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn
3543 ATACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAAT
     TATGTCTTCGTCCCCGTTCCGGTTACCTGTATAGTTTAAATAGTTCTCGGTAAATTTTTA

     3594 ahaIII,
```

FIG. 4G

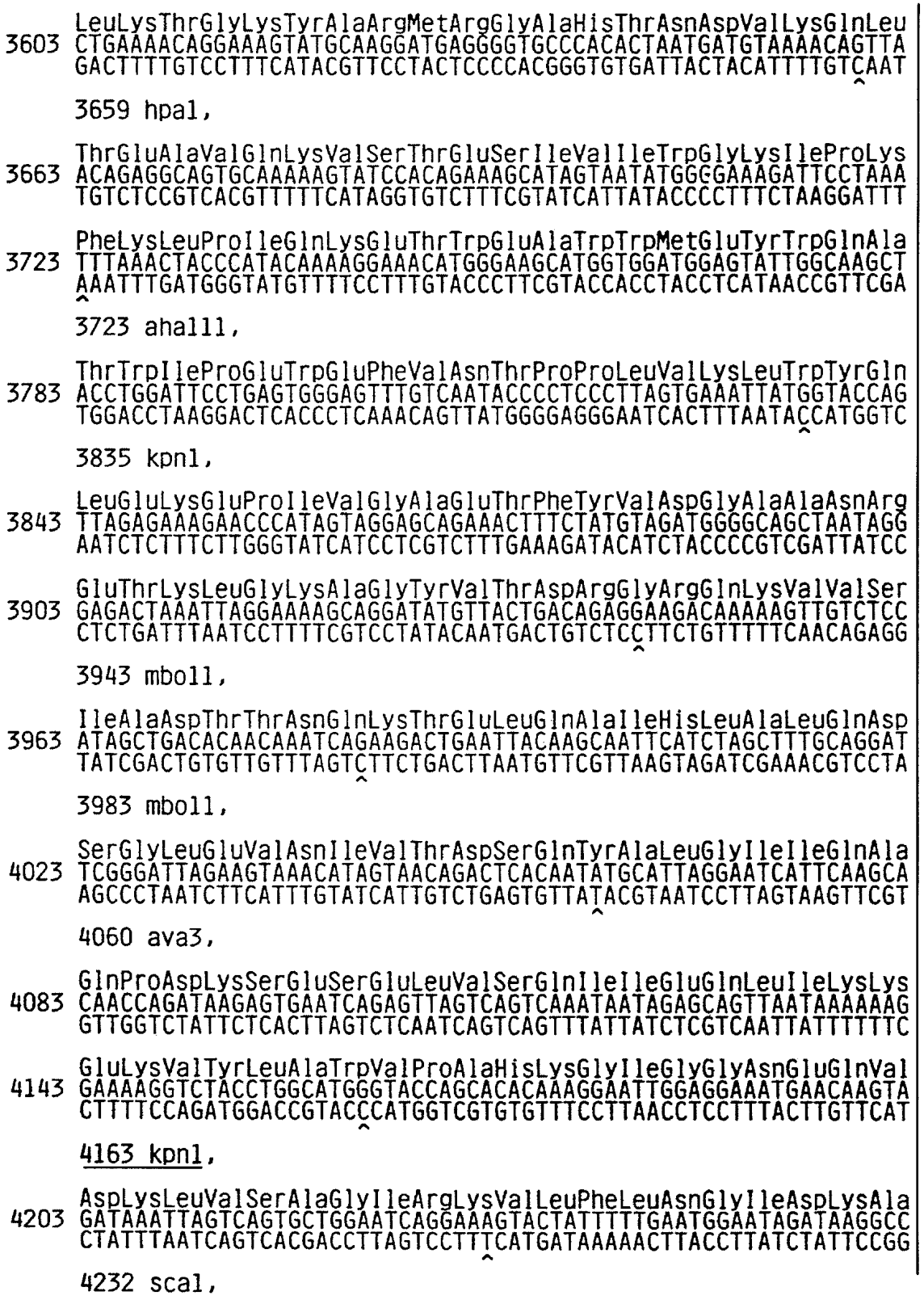

LeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeu
3603 CTGAAAACAGGAAAGTATGCAAGGATGAGGGGTGCCCACACTAATGATGTAAAACAGTTA
GACTTTTGTCCTTTCATACGTTCCTACTCCCCACGGGTGTGATTACTACATTTTGTCAAT
3659 hpa1,
ThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLysIleProLys
3663 ACAGAGGCAGTGCAAAAAGTATCCACAGAAAGCATAGTAATATGGGGAAAGATTCCTAAA
TGTCTCCGTCACGTTTTTCATAGGTGTCTTTCGTATCATTATACCCCTTTCTAAGGATTT
PheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAla
3723 TTTAAACTACCCATACAAAAGGAAACATGGGAAGCATGGTGGATGGAGTATTGGCAAGCT
AAATTTGATGGGTATGTTTTCCTTTGTACCCTTCGTACCACCTACCTCATAACCGTTCGA
3723 aha111,
ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln
3783 ACCTGGATTCCTGAGTGGGAGTTTGTCAATACCCCTCCCTTAGTGAAATTATGGTACCAG
TGGACCTAAGGACTCACCCTCAAACAGTTATGGGGAGGGAATCACTTTAATACCATGGTC
3835 kpn1,
LeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg
3843 TTAGAGAAAGAACCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCTAATAGG
AATCTCTTTCTTGGGTATCATCCTCGTCTTTGAAAGATACATCTACCCCGTCGATTATCC
GluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValValSer
3903 GAGACTAAATTAGGAAAAGCAGGATATGTTACTGACAGAGGAAGACAAAAAGTTGTCTCC
CTCTGATTTAATCCTTTTCGTCCTATACAATGACTGTCTCCTTCTGTTTTTCAACAGAGG
3943 mbo11,
IleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAsp
3963 ATAGCTGACACAACAAATCAGAAGACTGAATTACAAGCAATTCATCTAGCTTTGCAGGAT
TATCGACTGTGTTGTTTAGTCTTCTGACTTAATGTTCGTTAAGTAGATCGAAACGTCCTA
3983 mbo11,
SerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla
4023 TCGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCA
AGCCCTAATCTTCATTTGTATCATTGTCTGAGTGTTATACGTAATCCTTAGTAAGTTCGT
4060 ava3,
GlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIleLysLys
4083 CAACCAGATAAGAGTGAATCAGAGTTAGTCAGTCAAATAATAGAGCAGTTAATAAAAAAG
GTTGGTCTATTCTCACTTAGTCTCAATCAGTCAGTTTATTATCTCGTCAATTATTTTTTC
GluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal
4143 GAAAAGGTCTACCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTA
CTTTTCCAGATGGACCGTACCCATGGTCGTGTGTTTCCTTAACCTCCTTTACTTGTTCAT
4163 kpn1,
AspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIleAspLysAla
4203 GATAAATTAGTCAGTGCTGGAATCAGGAAAGTACTATTTTTGAATGGAATAGATAAGGCC
CTATTTAATCAGTCACGACCTTAGTCCTTTCATGATAAAAACTTACCTTATCTATTCCGG
4232 sca1,

FIG. 4H

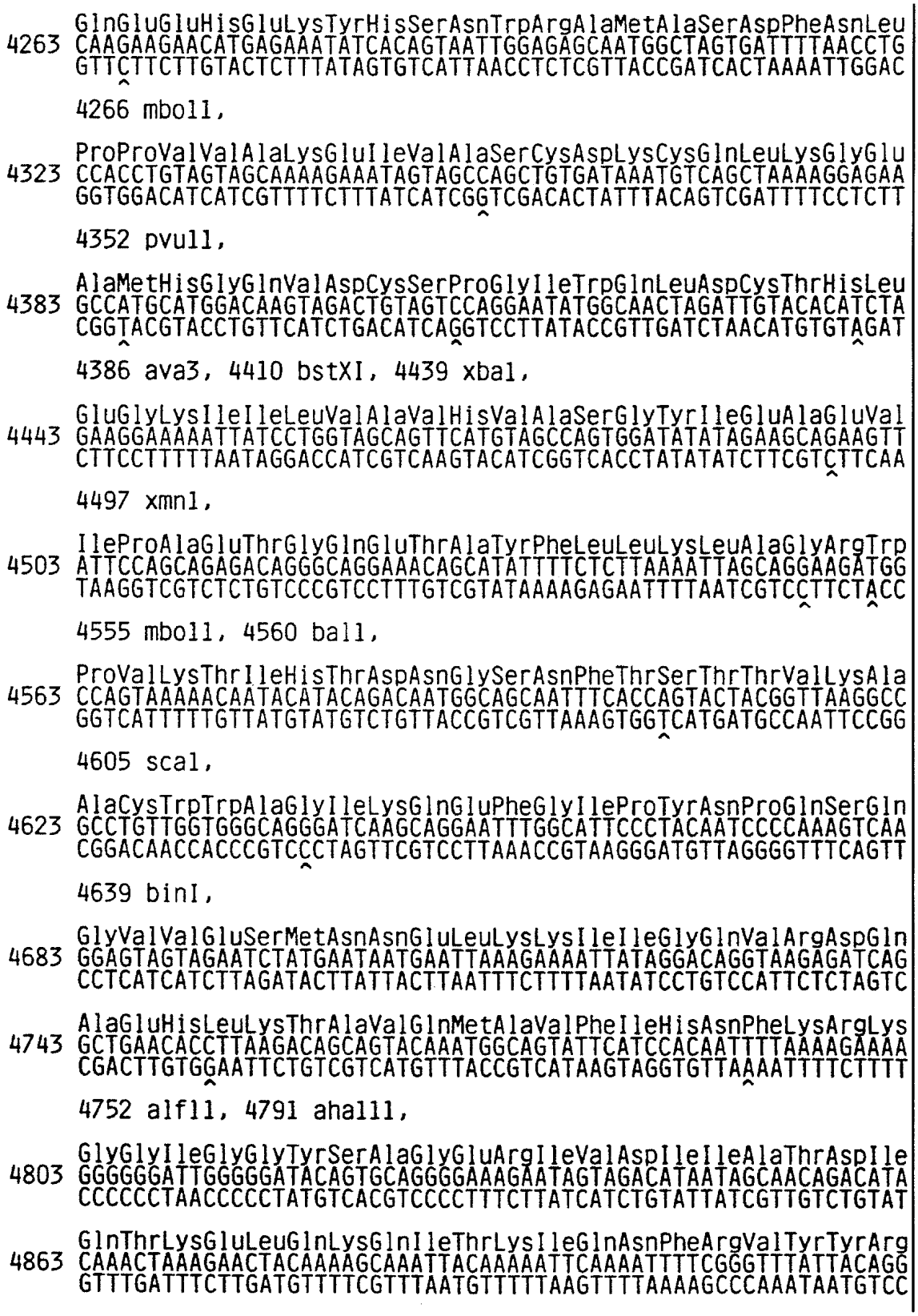

GlnGluGluHisGluLysTyrHisSerAsnTrpArgAlaMetAlaSerAspPheAsnLeu
4263 CAAGAAGAACATGAGAAATATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTG
GTTCTTCTTGTACTCTTTATAGTGTCATTAACCTCTCGTTACCGATCACTAAAATTGGAC
4266 mboll,
ProProValValAlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGlu
4323 CCACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAA
GGTGGACATCATCGTTTTCTTTATCATCGGTCGACACTATTTACAGTCGATTTTCCTCTT
4352 pvull,
AlaMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAspCysThrHisLeu
4383 GCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATCTA
CGGTACGTACCTGTTCATCTGACATCAGGTCCTTATACCGTTGATCTAACATGTGTAGAT
4386 ava3, 4410 bstXI, 4439 xbal,
GluGlyLysIleIleLeuValAlaValHisValAlaSerGlyTyrIleGluAlaGluVal
4443 GAAGGAAAAATTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTT
CTTCCTTTTTAATAGGACCATCGTCAAGTACATCGGTCACCTATATATCTTCGTCTTCAA
4497 xmnl,
IleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLeuLysLeuAlaGlyArgTrp
4503 ATTCCAGCAGAGACAGGGCAGGAAACAGCATATTTTCTCTTAAAATTAGCAGGAAGATGG
TAAGGTCGTCTCTGTCCCGTCCTTTGTCGTATAAAAGAGAATTTTAATCGTCCTTCTACC
4555 mboll, 4560 ball,
ProValLysThrIleHisThrAspAsnGlySerAsnPheThrSerThrThrValLysAla
4563 CCAGTAAAAACAATACATACAGACAATGGCAGCAATTTCACCAGTACTACGGTTAAGGCC
GGTCATTTTTGTTATGTATGTCTGTTACCGTCGTTAAAGTGGTCATGATGCCAATTCCGG
4605 scal,
AlaCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGln
4623 GCCTGTTGGTGGGCAGGGATCAAGCAGGAATTTGGCATTCCCTACAATCCCCAAAGTCAA
CGGACAACCACCCGTCCCTAGTTCGTCCTTAAACCGTAAGGGATGTTAGGGGTTTCAGTT
4639 binl,
GlyValValGluSerMetAsnAsnGluLeuLysLysIleIleGlyGlnValArgAspGln
4683 GGAGTAGTAGAATCTATGAATAATGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAG
CCTCATCATCTTAGATACTTATTACTTAATTTCTTTTAATATCCTGTCCATTCTCTAGTC
AlaGluHisLeuLysThrAlaValGlnMetAlaValPheIleHisAsnPheLysArgLys
4743 GCTGAACACCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAA
CGACTTGTGGAATTCTGTCGTCATGTTTACCGTCATAAGTAGGTGTTAAAATTTTCTTTT
4752 alfll, 4791 ahalll,
GlyGlyIleGlyGlyTyrSerAlaGlyGluArgIleValAspIleIleAlaThrAspIle
4803 GGGGGGATTGGGGGATACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATA
CCCCCCTAACCCCCTATGTCACGTCCCCTTTCTTATCATCTGTATTATCGTTGTCTGTAT
GlnThrLysGluLeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArg
4863 CAAACTAAAGAACTACAAAAGCAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGG
GTTTGATTTCTTGATGTTTTCGTTTAATGTTTTTAAGTTTTAAAAGCCCAAATAATGTCC

FIG. 4I

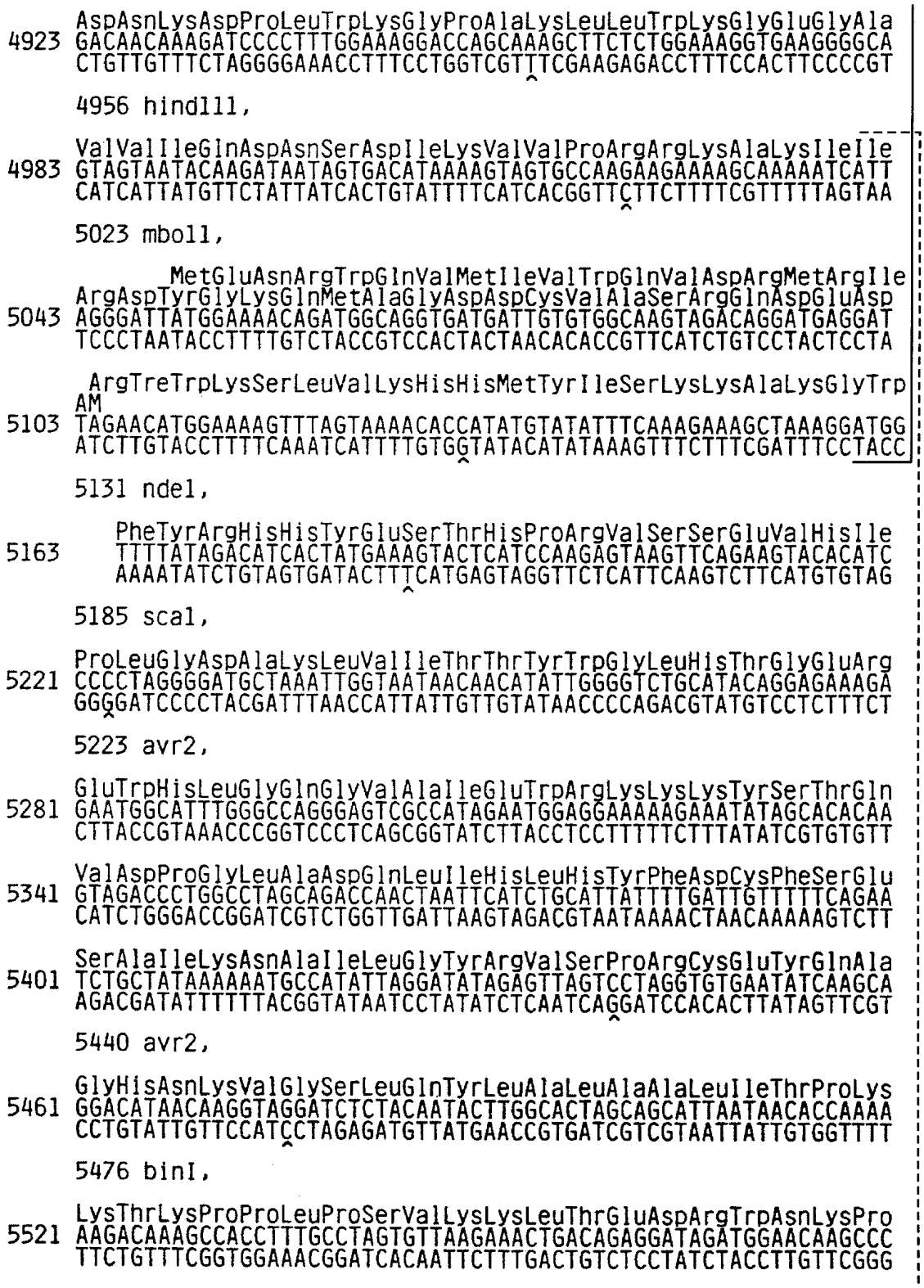

AspAsnLysAspProLeuTrpLysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAla
4923 GACAACAAAGATCCCCTTTGGAAAGGACCAGCAAAGCTTCTCTGGAAAGGTGAAGGGGCA
CTGTTGTTTCTAGGGGAAACCTTTCCTGGTCGTTTCGAAGAGACCTTTCCACTTCCCCGT
4956 hindIII,
ValValIleGlnAspAsnSerAspIleLysValValProArgArgLysAlaLysIleIle
4983 GTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAAGCAAAAATCATT
CATCATTATGTTCTATTATCACTGTATTTTCATCACGGTTCTTCTTTTCGTTTTTAGTAA
5023 mboII,
MetGluAsnArgTrpGlnValMetIleValTrpGlnValAspArgMetArgIle
ArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArgGlnAspGluAsp
5043 AGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGAT
TCCCTAATACCTTTTGTCTACCGTCCACTACTAACACACCGTTCATCTGTCCTACTCCTA
ArgTreTrpLysSerLeuValLysHisHisMetTyrIleSerLysLysAlaLysGlyTrp
AM
5103 TAGAACATGGAAAAGTTTAGTAAAACACCATATGTATATTTCAAAGAAAGCTAAAGGATGG
ATCTTGTACCTTTTCAAATCATTTTGTGGTATACATATAAAGTTTCTTTCGATTTCCTACC
5131 ndeI,
PheTyrArgHisHisTyrGluSerThrHisProArgValSerSerGluValHisIle
5163 TTTTATAGACATCACTATGAAAGTACTCATCCAAGAGTAAGTTCAGAAGTACACATC
AAAATATCTGTAGTGATACTTTCATGAGTAGGTTCTCATTCAAGTCTTCATGTGTAG
5185 scaI,
ProLeuGlyAspAlaLysLeuValIleThrThrTyrTrpGlyLeuHisThrGlyGluArg
5221 CCCCTAGGGGATGCTAAATTGGTAATAACAACATATTGGGGTCTGCATACAGGAGAAAGA
GGGGATCCCCTACGATTTAACCATTATTGTTGTATAACCCCAGACGTATGTCCTCTTTCT
5223 avrII,
GluTrpHisLeuGlyGlnGlyValAlaIleGluTrpArgLysLysLysTyrSerThrGln
5281 GAATGGCATTTGGGCCAGGGAGTCGCCATAGAATGGAGGAAAAAGAAATATAGCACACAA
CTTACCGTAAACCCGGTCCCTCAGCGGTATCTTACCTCCTTTTTCTTTATATCGTGTGTT
ValAspProGlyLeuAlaAspGlnLeuIleHisLeuHisTyrPheAspCysPheSerGlu
5341 GTAGACCCTGGCCTAGCAGACCAACTAATTCATCTGCATTATTTTGATTGTTTTTCAGAA
CATCTGGGACCGGATCGTCTGGTTGATTAAGTAGACGTAATAAAACTAACAAAAAGTCTT
SerAlaIleLysAsnAlaIleLeuGlyTyrArgValSerProArgCysGluTyrGlnAla
5401 TCTGCTATAAAAAATGCCATATTAGGATATAGAGTTAGTCCTAGGTGTGAATATCAAGCA
AGACGATATTTTTTACGGTATAATCCTATATCTCAATCAGGATCCACACTTATAGTTCGT
5440 avrII,
GlyHisAsnLysValGlySerLeuGlnTyrLeuAlaLeuAlaAlaLeuIleThrProLys
5461 GGACATAACAAGGTAGGATCTCTACAATACTTGGCACTAGCAGCATTAATAACACCAAAA
CCTGTATTGTTCCATCCTAGAGATGTTATGAACCGTGATCGTCGTAATTATTGTGGTTTT
5476 binI,
LysThrLysProProLeuProSerValLysLysLeuThrGluAspArgTrpAsnLysPro
5521 AAGACAAAGCCACCTTTGCCTAGTGTTAAGAAACTGACAGAGGATAGATGGAACAAGCCC
TTCTGTTTCGGTGGAAACGGATCACAATTCTTTGACTGTCTCCTATCTACCTTGTTCGGG

FIG. 4J

```
    GlnLysThrLysGlyHisArgGlySerHisThrMetAsnGlyHisAM
5581 CAGAAGACCAAGGGCCACAGAGGGAGCCATACAATGAATGGACACTAGAGCTTTTAGAGG
     GTCTTCTGGTTCCCGGTGTCTCCCTCGGTATGTTACTTACCTGTGATCTCGAAAATCTCC

     5583 mboII,

5641 AGCTTAAGAGAGAAGCTGTTAGACATTTTCCTAGGCCATGGCTCCATAGCTTAGGACAAT
     TCGAATTCTCTCTTCGACAATCTGTAAAAGGATCCGGTACCGAGGTATCGAATCCTGTTA

     5643 afIII, 5670 avr2, 5676 ncoI,

5701 ATATCTATGAAACTTATGGGGATACTTGGGCAGGAGTGGAAGCCATAATAAGAATTCTGC
     TATAGATACTTTGAATACCCCTATGAACCCGTCCTCACCTTCGGTATTATTCTTAAGACG

     5752 ecorI,

5761 AACAACTGCTGTTTATTCATTTCAGAATTGGGTGTCAACATAGCAGAATAGGCATTATTC
     TTGTTGACGACAAATAAGTAAAGTCTTAACCCACAGTTGTATCGTCTTATCCGTAATAAG

5821 AACAGAGGAGAGCAAGAAGAAATGGAGCCAGTAGATCCTAATCTAGAGCCCTGGAAGCAT
     TTGTCTCCTCTCGTTCTTCTTTACCTCGGTCATCTAGGATTAGATCTCGGGACCTTCGTA

     5836 mboII, 5862 xbaI,

5881 CCAGGAAGTCAGCCTAGGACTGCTTGTAACAATTGCTATTGTAAAAAGTGTTGCTTTCAT
     GGTCCTTCAGTCGGATCCTGACGAACATTGTTAACGATAACATTTTTCACAACGAAAGTA

     5893 avr2,

5941 TGCTACGCGTGTTTCACAAGAAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGA
     ACGATGCGCACAAAGTGTTCTTTTCCGAATCCGTAGAGGATACCGTCCTTCTTCGCCTCT

     5945 mluI, 5988 mboII,

6001 CAGCGACGAAGAGCTCCTCAGGACAGTCAGACTCATCAAGCTTCTCTATCAAAGCAGTAA
     GTCGCTGCTTCTCGAGGAGTCCTGTCAGTCTGAGTAGTTCGAAGAGATAGTTTCGTCATT

     6008 mboII, 6011 sacI, 6016 mstII, 6038 hindIII,

6061 GTAGTAAATGTAATGCAATCTTTACAAATATTAGCAATAGTATCATTAGTAGTAGTAGCA
     CATCATTTACATTACGTTAGAAATGTTTATAATCGTTATCATAGTAATCATCATCATCGT

6121 ATAATAGCAATAGTTGTGTGGACCATAGTACTCATAGAATATAGGAAAATATTAAGACAA
     TATTATCGTTATCAACACACCTGGTATCATGAGTATCTTATATCCTTTTATAATTCTGTT

     6147 scaI,

                                                          MetLys  ENV
6181 AGAAAATAGACAGATTAATTGATAGAATAAGAGAAAAAGCAGAAGACAGTGGCAATGAAA
     TCTTTTATCTGTCTAATTAACTATCTTATTCTCTTTTTCGTCTTCTGTCACCGTTACTTT

     6222 mboII,

     ValLysGlyThrArgArgAsnTyrGlnHisLeuTrpArgTrpGlyThrLeuLeuLeuGly
6241 GTGAAGGGGACCAGGAGGAATTATCAGCACTTGTGGAGATGGGGCACCTTGCTCCTTGGG
     CACTTCCCCTGGTCCTCCTTAATAGTCGTGAACACCTCTACCCCGTGGAACGAGGAACCC

     MetLeuMetIleCysSerAlaThrGluLysLeuTrpValThrValTyrTyrGlyValPro
6301 ATGTTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTTTATTATGGAGTACCT
     TACAACTACTAGACATCACGATGTCTTTTTAACACCCAGTGTCAAATAATACCTCATGGA
```

FIG. 4K

```
      ValTrpLysGluAlaThrThrThrLeuPheCysAlaSerAspAlaArgAlaTyrAspThr
6361  GTGTGGAAAGAAGCAACTACCACTCTATTTTGTGCATCAGATGCTAGAGCATATGATACA
      CACACCTTTCTTCGTTGATGGTGAGATAAAACACGTAGTCTACGATCTCGTATACTATGT

      6410 ndeI,

      GluValHisAsnValTrpAlaThrHisAlaCysValProThrAspProAsnProGlnGlu
6421  GAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAA
      CTCCATGTATTACAAACCCGGTGTGTACGGACACATGGGTGTCTGGGGTTGGGTGTTCTT

      ValValLeuGlyAsnValThrGluAsnPheAsnMetTrpLysAsnAsnMetValGluGln
6481  GTAGTATTGGGAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACATGGTAGAACAG
      CATCATAACCCTTTACACTGTCTTTTAAAATTGTACACCTTTTTATTGTACCATCTTGTC

      MetGlnGluAspIleIleSerLeuTrpAspGlnSerLeuLysProCysValLysLeuThr
6541  ATGCAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACC
      TACGTCCTCCTATATTAGTCAAATACCCTAGTTTCGGATTTCGGTACACATTTTAATTGG

      6567 binI,

      ProLeuCysValThrLeuAsnCysThrAspLeuGlyLysAlaThrAsnThrAsnSerSer
6601  CCACTCTGTGTTACTTTAAATTGCACTGATTTGGGGAAGGCTACTAATACCAATAGTAGT
      GGTGAGACACAATGAAATTTAACGTGACTAAACCCCTTCCGATGATTATGGTTATCATCA

      6615 ahaIII,

      AsnTrpLysGluGluIleLysGlyGluIleLysAsnCysSerPheAsnIleThrThrSer
6661  AATTGGAAAGAAGAAATAAAAGGAGAAATAAAAAACTGCTCTTTCAATATCACCACAAGC
      TTAACCTTTCTTCTTTATTTTCCTCTTTATTTTTTGACGAGAAAGTTATAGTGGTGTTCG

      6670 mboII,

      IleArgAspLysIleGlnLysGluAsnAlaLeuPheArgAsnLeuAspValValProIle
6721  ATAAGAGATAAGATTCAGAAAGAAAATGCACTTTTTCGTAACCTTGATGTAGTACCAATA
      TATTCTCTATTCTAAGTCTTTCTTTTACGTGAAAAAGCATTGGAACTACATCATGGTTAT

      AspAsnAlaSerThrThrThrAsnTyrThrAsnTyrArgLeuIleHisCysAsnArgSer
6781  GATAATGCTAGTACTACTACCAACTATACCAACTATAGGTTGATACATTGTAACAGATCA
      CTATTACGATCATGATGATGGTTGATATGGTTGATATCCAACTATGTAACATTGTCTAGT

      6790 scaI,

      ValIleThrGlnAlaCysProLysValSerPheGluProIleProIleHisTyrCysThr
6841  GTCATTACACAGGCCTGTCCAAAGGTATCATTTGAGCCAATTCCCATACATTATTGTACC
      CAGTAATGTGTCCGGACAGGTTTCCATAGTAAACTCGGTTAAGGGTATGTAATAACATGG

      6851 stuI,

      ProAlaGlyPheAlaIleLeuLysCysAsnAsnLysThrPheAsnGlyLysGlyProCys
6901  CCGGCTGGTTTTGCGATTCTAAAGTGTAATAATAAAACGTTCAATGGAAAAGGACCATGT
      GGCCGACCAAAACGCTAAGATTTCACATTATTATTTTGCAAGTTACCTTTTCCTGGTACA

      ThrAsnValSerThrValGlnCysThrHisGlyIleArgProIleValSerThrGlnLeu
6961  ACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAATAGTGTCAACTCAACTG
      TGTTTACAGTCGTGTCATGTTACATGTGTACCTTAATCCGGTTATCACAGTTGAGTTGAC

      LeuLeuAsnGlySerLeuAlaGluGluGluValValIleArgSerAspAsnPheThrAsn
7021  CTGTTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGACAATTTCACGAAC
      GACAATTTACCGTCAGATCGTCTTCTTCTCCATCATTAATCTAGACTGTTAAAGTGCTTG

      7042 mboII, 7045 mboII, 7060 bglII,
```

FIG. 4L

```
    AsnAlaLysThrIleIleValGlnLeuAsnGluSerValAlaIleAsnCysThrArgPro
7081 AATGCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATTAACTGTACAAGACCC
     TTACGATTTTGGTATTATCATGTCGACTTACTTAGACATCGTTAATTGACATGTTCTGGG

     7102 pvu11,

    AsnAsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAlaPheHisThrThrGly
7141 AACAACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCATTTCATACAACAGGA
     TTGTTGTTATGTTCTTTTTCATAGATATATCCTGGTCCCTCTCGTAAAGTATGTTGTCCT

     7199 mbo11,

    ArgIleIleGlyAspIleArgLysAlaHisCysAsnIleSerArgAlaGlnTrpAsnAsn
7201 AGAATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGAGCACAATGGAATAAC
     TCTTATTATCCTCTATATTCTTTTCGTGTAACATTGTAATCATCTCGTGTTACCTTATTG

    ThrLeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsnAsnLysThrIleVal
7261 ACTTTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAATAATAAAACAATAGTC
     TGAAATCTTGTCTATCAATTTTTTAATTCTCTTGTCAAACCCTTATTATTTTGTTATCAG

    PheAsnGlnSerSerGlyGlyAspProGluIleValMetHisSerPheAsnCysArgGly
7321 TTTAATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGG
     AAATTAGTTAGGAGTCCTCCCCTGGGTCTTTAACATTACGTGTCAAAATTAACATCTCCC

     7331 mstII,

    GluPhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrpArgLeuAsnHisThr
7381 GAATTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGGAGGTTAAATCACACT
     CTTAAAAAAGATGACATTATGTTGTGTTGACAAATTATTATGTACCTCCAATTTAGTGTGA

    GluGlyThrLysGlyAsnAspThrIleIleLeuProCysArgIleLysGlnIleIleAsn
7441 GAAGGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATAAAACAAATTATAAAC
     CTTCCTTGATTTCCTTTACTGTGTTAGTATGAGGGTACATCTTATTTTGTTTAATATTTG

    MetTrpGlnGluValGlyLysAlaMetTyrAlaProProIleGlyGlyGlnIleSerCys
7501 ATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGAGGACAAATTAGTTGT
     TACACCGTCCTTCATCCTTTTCGTTACATACGGGGAGGGTAACCTCCTGTTTAATCAACA

    SerSerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThrAsnValThrAsnAsp
7561 TCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACAAATGTAACTAATGAC
     AGTAGTTTATAATGTCCCGACGATAATTGTTCTCTACCACCATGTTTACATTGATTACTG

    ThrGluValPheArgProGlyGlyGlyAspMetArgAspAsnTrpArgSerGluLeuTyr
7621 ACCGAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATAT
     TGGCTCCAGAAGTCTGGACCTCCTCCTCTATACTCCCTGTTAACCTCTTCACTTAATATA

     7628 mbo11,

    LysTyrLysValIleLysIleGluProLeuGlyIleAlaProThrLysAlaLysArgArg
7681 AAATATAAAGTAATAAAAATTGAACCATTAGGAATAGCACCCACCAAGGCAAAGAGAAGA
     TTTATATTTCATTATTTTTAACTTGGTAATCCTTATCGTGGGTGGTTCCGTTTCTCTTCT

     7736 mbo11,

    ValValGlnArgGluLysArgAlaValGlyIleValGlyAlaMetPheLeuGlyPheLeu
7741 GTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGTAGGAGCTATGTTCCTTGGGTTCTTG
     CACCACGTCTCTCTTTTTTCTCGTCACCCTTATCATCCTCGATACAAGGAACCCAAGAAC

    GlyAlaAlaGlySerThrMetGlyAlaValSerLeuThrLeuThrValGlnAlaArgGln
7801 GGAGCAGCAGGAAGCACTATGGGCGCAGTGTCATTGACGCTGACGGTACAGGCCAGACAA
     CCTCGTCGTCCTTCGTGATACCCGCGTCACAGTAACTGCGACTGCCATGTCCGGTCTGTT
```

FIG. 4M

```
      LeuLeuSerGlyIleValGlnGlnGlnAsnAsnLeuLeuArgAlaIleGluAlaGlnGln
7861  TTATTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAA
      AATAACAGACCATATCACGTTGTCGTCTTGTTAAACGACTCCCGATAACTCCGCGTTGTT

      HisLeuLeuGlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaVal
7921  CATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGTG
      GTAGACAACGTTGAGTGTCAGACCCCGTAGTTCGTCGAGGTCCGTTCTCAGGACCGACAC

      GluArgTyrLeuArgAspGlnGlnLeuLeuGlyIleTrpGlyCysSerGlyLysLeuIle
7981  GAAAGATACCTAAGGGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTGGAAAACTCATT
      CTTTCTATGGATTCCCTAGTTGTCGAGGATCCCTAAACCCCAACGAGACCTTTTGAGTAA

      7989 mstII, 7995 binI, 8007 avr2,

      CysThrThrAlaValProTrpAsnAlaSerTrpSerAsnLysSerLeuGluAspIleTrp
8041  TGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAAGACATTTGG
      ACGTGGTGACGACACGGAACCTTACGATCAACCTCATTATTTAGAGACCTTCTGTAAACC

      8089 mboII,

      AspAsnMetThrTrpMetGlnTrpGluArgGluIleAspAsnTyrThrAsnThrIleTyr
8101  GATAACATGACCTGGATGCAGTGGGAAAGAGAAATTGACAATTACACAAACACAATATAC
      CTATTGTACTGGACCTACGTCACCCTTTCTCTTTAACTGTTAATGTGTTTGTGTTATATG

      ThrLeuLeuGluGluSerGlnAsnGlnGlnGluLysAsnGluGlnGluLeuLeuGluLeu
8161  ACCTTACTTGAAGAATCGCAGAACCAACAAGAAAAGAATGAACAAGAATTATTAGAATTG
      TGGAATGAACTTCTTAGCGTCTTGGTTGTTCTTTTCTTACTTGTTCTTAATAATCTTAAC

      8170 mboII,

      AspLysTrpAlaSerLeuTrpAsnTrpPheSerIleThrAsnTrpLeuTrpTyrIleLys
8221  GATAAGTGGGCAAGTTTGTGGAATTGGTTTAGCATAACAAACTGGCTGTGGTATATAAAG
      CTATTCACCCGTTCAAACACCTTAACCAAATCGTATTGTTTGACCGACACCATATATTTC

      IlePheIleMetIleValGlyGlyLeuValGlyLeuArgIleValPheAlaValLeuSer
8281  ATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTGCTTTCT
      TATAAGTATTACTATCATCCTCCGAACCATCCAAATTCTTATCAAAAACGACACGAAAGA

      IleValAsnArgValArgGlnGlyTyrSerProLeuSerPheGlnThrArgLeuProVal
8341  ATAGTGAATAGAGTTAGGCAGGGATACTCACCATTGTCATTTCAGACCCGCCTCCCAGTC
      TATCACTTATCTCAATCCGTCCCTATGAGTGGTAACAGTAAAGTCTGGGCGGAGGGTCAG

      8400 aval,

      ProArgGlyProAspArgProAspGlyIleGluGluGluGlyGlyGluArgAspArgAsp
8401  CCGAGGGGACCCGACAGGCCCGACGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGAGAC
      GGCTCCCCTGGGCTGTCCGGGCTGCCTTAGCTTCTTCTTCCACCTCTCTCTCTGTCTCTG

      8431 mboII, 8434 mboII,

      ArgSerValArgLeuValAspGlyPheLeuAlaLeuIleTrpGluAspLeuArgSerLeu
8461  AGATCCGTTCGATTAGTGGATGGATTCTTAGCACTTATCTGGGAAGATCTGCGGAGCCTG
      TCTAGGCAAGCTAATCACCTACCTAAGAATCGTGAATAGACCCTTCTAGACGCCTCGGAC

      8503 mboII, 8505 bglII,

      CysLeuPheSerTyrArgArgLeuArgAspLeuLeuLeuIleAlaAlaArgThrValGlu
8521  TGCCTCTTCAGCTACCGCCGCTTGAGAGACTTACTCTTGATTGCAGCGAGGACTGTGGAA
      ACGGAGAAGTCGATGGCGGCGAACTCTCTGAATGAGAACTAACGTCGCTCCTGACACCTT

      8525 mboII,
```

FIG. 4N

```
                IleLeuGlyHisArgGlyTrpGluAlaLeuLysTyrTrpTrpSerLeuLeuGlnTyrTrp
8581            ATTCTGGGGCACAGGGGGTGGGAAGCCCTCAAATATTGGTGGAGTCTCCTGCAGTATTGG
                TAAGACCCCGTGTCCCCCACCCTTCGGGAGTTTATAACCACCTCAGAGGACGTCATAACC

                8629 pst1,

                IleGlnGluLeuLysAsnSerAlaValSerTrpLeuAsnAlaThrAlaIleAlaValThr
8641            ATTCAGGAACTAAAGAATAGTGCTGTTAGCTGGCTCAACGCCACAGCTATAGCAGTAACT
                TAAGTCCTTGATTTCTTATCACGACAATCGACCGAGTTGCGGTGTCGATATCGTCATTGA

                GluGlyThrAspArgValIleGluValAlaGlnArgAlaTyrArgAlaIleLeuHisIle
8701            GAGGGGACAGATAGGGTTATAGAAGTAGCACAAAGAGCTTATAGAGCTATTCTCCACATA
                CTCCCCTGTCTATCCCAATATCTTCATCGTGTTTCTCGAATATCTCGATAAGAGGTGTAT

                HisArgArgIleArgGlnGlyLeuGluArgLeuLeuLeuOC  MetGlyGlyLysTrpSer
8761            CATAGAAGAATTAGACAGGGCTTGGAAAGGCTTTTGCTATAAGATGGGTGGCAAGTGGTCA
                GTATCTTCTTAATCTGTCCCGAACCTTTCCGAAAACGATATTCTACCCACCGTTCACCAGT

                8765 mboll,

                LysArgSerMetGlyGlyTrpSerAlaIleArgGluArgMetArgArgAlaGluProArg
8822            AAACGTAGTATGGGTGGATGGTCTGCTATAAGGGAAAGAATGAGACGAGCTGAGCCACGA
                TTTGCATCATACCCACCTACCAGACGATATTCCCTTTCTTACTCTGCTCGACTCGGTGCT

                AlaGluProAlaAlaAspGlyValGlyAlaValSerArgAspLeuGluLysHisGlyAla
8882            GCTGAGCCAGCAGCAGATGGGGTGGGAGCAGTATCTCGAGACCTGGAAAAACATGGAGCA
                CGACTCGGTCGTCGTCTACCCCACCCTCGTCATAGAGCTCTGGACCTTTTTGTACCTCGT

                8883 tthIII1, 8916 aval xhol,

                IleThrSerSerAsnThrAlaAlaThrAsnAlaAspCysAlaTrpLeuGluAlaGlnGlu
8942            ATCACAAGTAGCAATACAGCAGCTACTAATGCTGATTGTGCCTGGCTAGAAGCACAAGAG
                TAGTGTTCATCGTTATGTCGTCGATGATTACGACTAACACGGACCGATCTTCGTGTTCTC

                GluGluGluValGlyPheProValArgProGlnValProLeuArgProMetThrTyrLys
9002            GAGGAAGAGGTGGGTTTTCCAGTCAGACCTCAGGTACCTTTAAGACCAATGACTTACAAG
                CTCCTTCTCCACCCAAAAGGTCAGTCTGGAGTCCATGGAAATTCTGGTTACTGAATGTTC

                9005 mboll, 9029 mstII, 9034 kpn1,

                AlaAlaLeuAspIleSerHisPheLeuLysGluLysGlyGlyLeuGluGlyLeuIleTrp
9062            GCAGCTTTAGATATTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTTGG
                CGTCGAAATCTATAATCGGTGAAAAATTTTCTTTTCCCCCCTGACCTTCCCGATTAAACC

                9085 aha111,

                SerGlnArgArgGlnGluIleLeuAspLeuTrpIleTyrHisThrGlnGlyTyrPhePro
9122            TCCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCT
                AGGGTTTCTTCTGTTCTCTAGGAACTAGACACCTAGATGGTGTGTGTTCCGATGAAGGGA

                9129 mboll, 9153 binI,

                AspTrpGlnAsnTyrThrProGlyProGlyIleArgTyrProLeuThrPheGlyTrpCys
9182            GATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGC
                CTAACCGTCTTAATGTGTGGTCCCGGTCCCTAGTCTATAGGTGACTGGAAACCTACCACG

                9210 binI, 9216 ecor5,
```

FIG. 4O

```
                PheLysLeuValProValGluProGluLysValGluGluAlaAsnGluGlyGluAsnAsn
9242  TTCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCCAATGAAGGAGAGAACAAC
      AAGTTCGATCATGGTCAACTCGGTCTCTTCCATCTTCTCCGGTTACTTCCTCTCTTGTTG

      9275 mboll,

                SerLeuLeuHisProMetSerLeuHisGlyMetGluAspAlaGluLysGluValLeuVal
9302  AGCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCGGAGAAAGAAGTGTTAGTG
      TCGAACAATGTGGGATACTCGGACGTACCCTACCTCCTGCGCCTCTTTCTTCACAATCAC

                TrpArgPheAspSerLysLeuAlaPheHisHisMetAlaArgGluLeuHisProGluTyr
9362  TGGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGAGTAC
      ACCTCCAAACTGTCGTTTGATCGTAAAGTAGTGTACCGGGCTCTCGACGTAGGCCTCATG

      9399 aval, 9417 scal,

                TyrLysAspCysOP
9422  TACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGG
      ATGTTTCTGACGACTGTAGCTCGAAAGATGTTCCCTGAAAGGCGACCCCTGAAAGGTCCC

9482  AGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATGCTGCATATAAGCAGCTG
      TCCGCACCGGACCCGCCCTGACCCCTCACCGCAGGGAGTCTACGACGTATATTCGTCGAC

      9536 pvull,

9542  CTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGC
      GAAAAACGGACATGACCCAGAGAGACCAATCTGGTCTAGACTCGGACCCTCGAGAGACCG

      9576 bglll, 9590 sacl,

9602  TAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTG
      ATTGATCCCTTGGGTGACGAATTCGGAGTTATTTCGAACGGAACTCACGAAGTTCATCAC

      9620 aflll, 9634 hindlll,

9662  TGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTG
      ACACGGGCAGACAACACACTGAGACCATTGATCTCTAGGGAGTCTGGGAAAATCAGTCAC

9722  TGGAAAAATCTCTAGCAG
      ACCTTTTTAGAGATCGTC
```

FIG. 4P

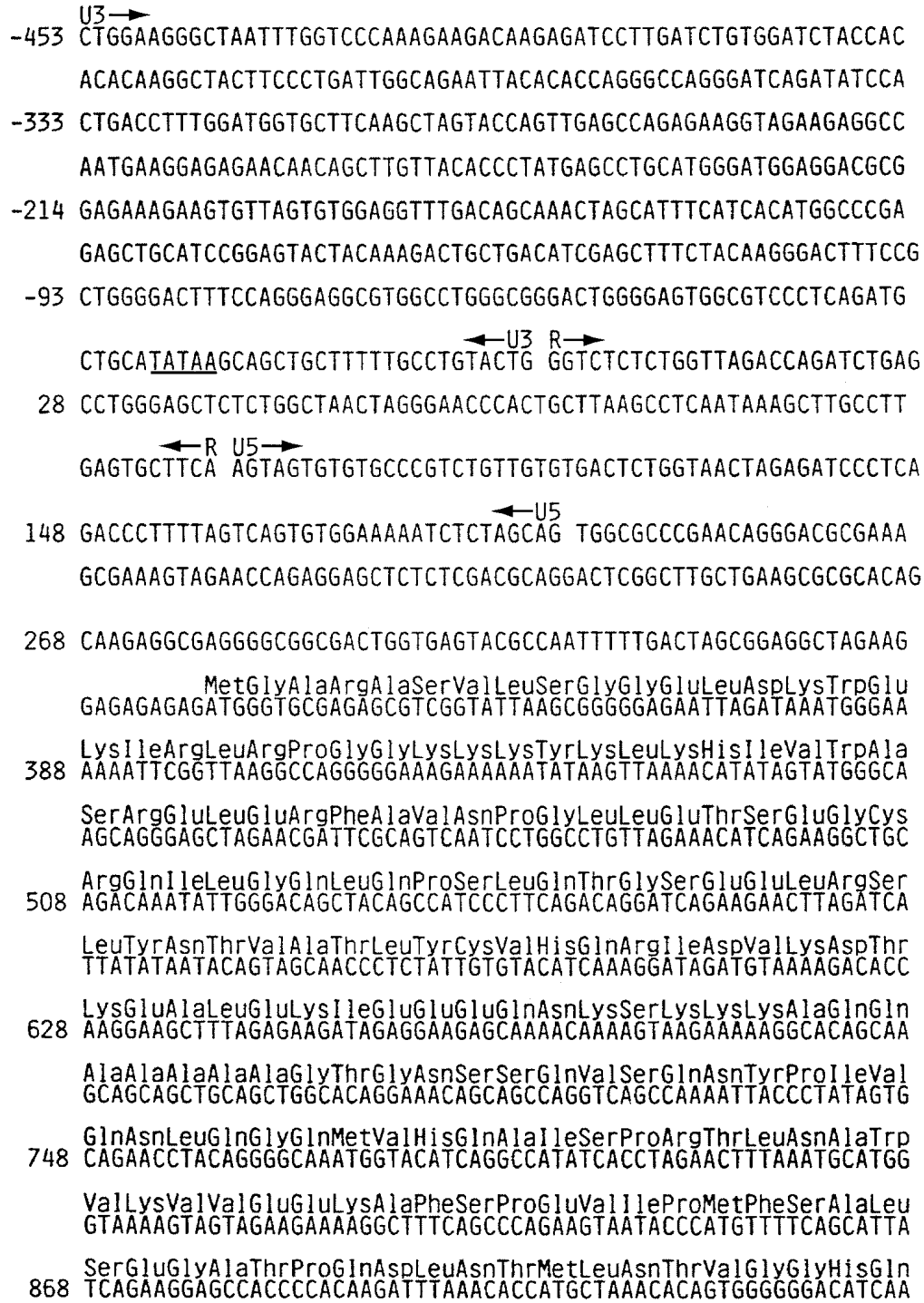

U3→
-453 CTGGAAGGGCTAATTTGGTCCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCAC
ACACAAGGCTACTTCCCTGATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCA
-333 CTGACCTTTGGATGGTGCTTCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCC
AATGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCG
-214 GAGAAAGAAGTGTTAGTGTGGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGA
GAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCG
-93 CTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATG
←U3 R→
CTGCATATAAGCAGCTGCTTTTTGCCTGTACTG GGTCTCTCTGGTTAGACCAGATCTGAG
28 CCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTT
←R U5→
GAGTGCTTCA AGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCA
←U5
148 GACCCTTTTAGTCAGTGTGGAAAAATCTCTAGCAG TGGCGCCCGAACAGGGACGCGAAA
GCGAAAGTAGAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACAG
268 CAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAATTTTTGACTAGCGGAGGCTAGAAG
MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspLysTrpGlu 17
GAGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAATGGGAA
LysIleArgLeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAla
388 AAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAGTTAAAACATATAGTATGGGCA
SerArgGluLeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCys 57
AGCAGGGAGCTAGAACGATTCGCAGTCAATCCTGGCCTGTTAGAAACATCAGAAGGCTGC
ArgGlnIleLeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSer
508 AGACAAATATTGGGACAGCTACAGCCATCCCTTCAGACAGGATCAGAAGAACTTAGATCA
LeuTyrAsnThrValAlaThrLeuTyrCysValHisGlnArgIleAspValLysAspThr 97
TTATATAATACAGTAGCAACCCTCTATTGTGTACATCAAAGGATAGATGTAAAAGACACC
LysGluAlaLeuGluLysIleGluGluGluGlnAsnLysSerLysLysLysAlaGlnGln
628 AAGGAAGCTTTAGAGAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAGGCACAGCAA
AlaAlaAlaAlaAlaGlyThrGlyAsnSerSerGlnValSerGlnAsnTyrProIleVal 137
GCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTCAGCCAAAATTACCCTATAGTG
GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrp
748 CAGAACCTACAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG
ValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu 177
GTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA
SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGln
868 TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAA
L
T
R

FIG. 5A

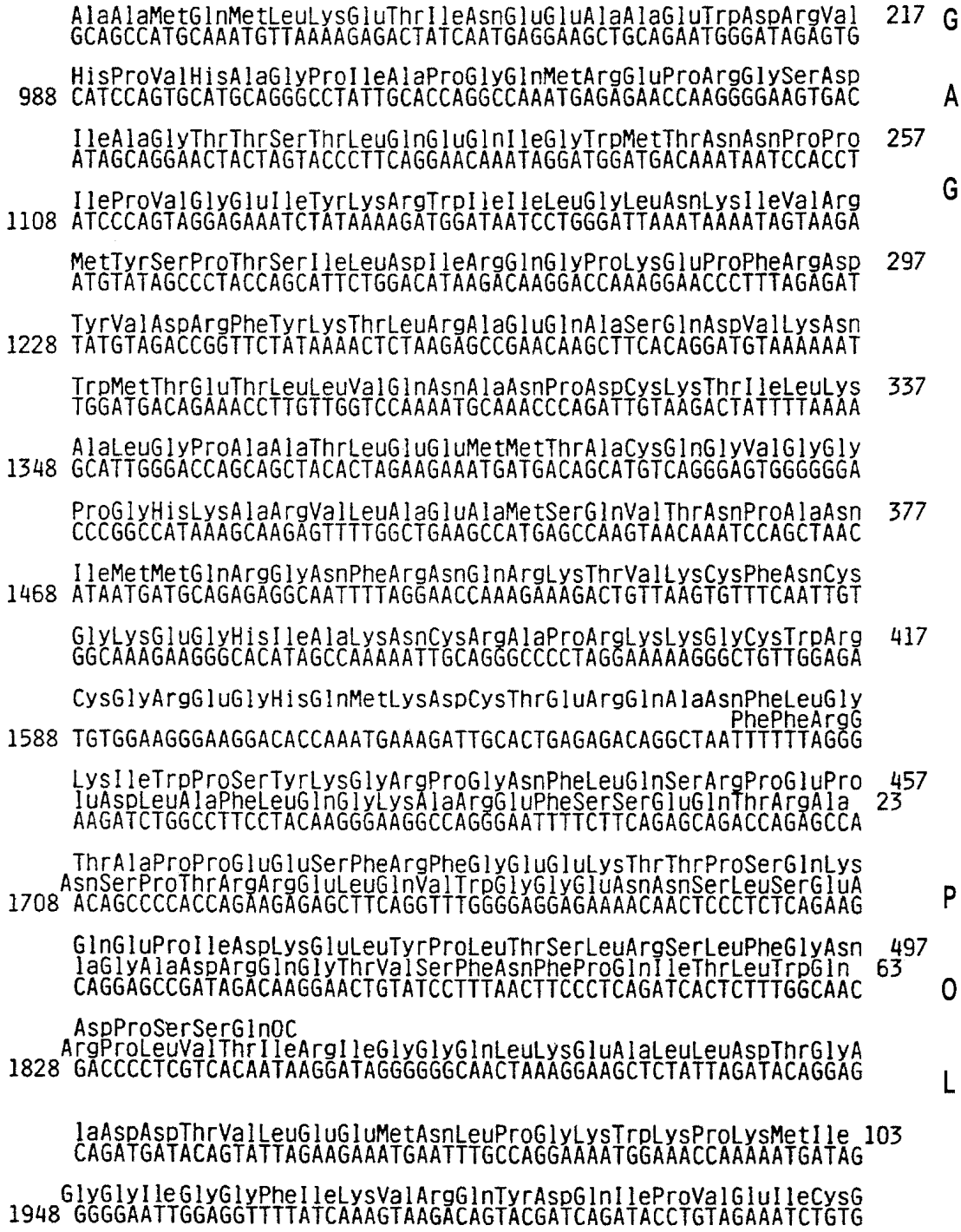

AlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal 217
GCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG
HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAsp
988 CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGAC
IleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProPro 257
ATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT
IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArg
1108 ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGA
MetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp 297
ATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT
TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsn
1228 TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAAT
TrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys 337
TGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA
AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGly
1348 GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGA
ProGlyHisLysAlaArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaAsn 377
CCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCCATGAGCCAAGTAACAAATCCAGCTAAC
IleMetMetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCys
1468 ATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGACTGTTAAGTGTTTCAATTGT
GlyLysGluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpArg 417
GGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAGA
CysGlyArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGly
PhePheArgG
1588 TGTGGAAGGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGG
LysIleTrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluPro 457
luAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAla 23
AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA
ThrAlaProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLys
AsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSerLeuSerGluA
1708 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG
GlnGluProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsn 497
laGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGln 63
CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC
AspProSerSerGlnOC
ArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyA
1828 GACCCCTCGTCACAATAAGGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAG
laAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLysMetIle 103
CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG
GlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProValGluIleCysG
1948 GGGGAATTGGAGGTTTTATCAAAGTAAGACAGTACGATCAGATACCTGTAGAAATCTGTG

FIG. 5B

```
     lyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArg 143
     GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA

    AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValP
2068 ATCTGTTGACTCAGATTGGTTGTACTTTAAATTTCCCCATTAGTCCTATTGAAACTGTAC

     roValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 183
     CAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAGCAATGGCCATTGACAGAAG

    GluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerL
2188 AAAAAATAAAAGCATTAGTAGAGATATGTACAGAAATGGAAAAGGAAGGGAAAATTTCAA

     ysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 223
     AAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCTATAAAGAAAAAAGACAGTA

    ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpG
2308 CTAAATGGAGAAAACTAGTAGATTTCAGAGAACTTAATAAAAGAACTCAAGACTTCTGGG

     luValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVal 263
     AAGTTCAGTTAGGAATACCACACCCCGCAGGGTTAAAAAAGAAAAAATCAGTAACAGTAT

    LeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLysTyrThrA
2428 TGGATGTGGGTGATGCATACTTTTCAGTTCCCTTAGATAAAGACTTTAGAAAGTATACTG

     laPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 303
     CATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGC

    LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuG
2548 TGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAG

     luProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyr 343
     AGCCTTTTAGAAAACAGAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATG

    ValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisL
2668 TAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAACTGAGACAGCATC

     euLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 383
     TGTTGAGGTGGGGATTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTT

    TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMetLeuProGluL
2788 GGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAATGCTGCCAGAAA

     ysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 423
     AAGACAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGAAAATTGAATTGGGCAAGTC

    GlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThrLysAlaL
2908 AGATTTATGCAGGGATTAAAGTAAAGCAGTTATGTAAACTCCTTAGAGGAACCAAAGCAC

     euThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 463  P
     TAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGGGAGA

    IleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeuValAlaGluI
3028 TTCTAAAAGAACCAGTACATGAAGTATATTATGACCCATCAAAAGACTTAGTAGCAGAAA

     leGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 503  O
     TACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAATC

    LeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuT
3148 TGAAAACAGGAAAGTATGCAAGGATGAGGGGTGCCCACACTAATGATGTAAAACAGTTAA

     hrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLysIleProLys 543  L
     CAGAGGCAGTGCAAAAAGTATCCACAGAAAGCATAGTAATATGGGGAAAGATTCCTAAAT
```

FIG. 5C

```
    PheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaT
3268 TTAAACTACCCATACAAAAGGAAACATGGGAAGCATGGTGGATGGAGTATTGGCAAGCTA

     hrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 583
     CCTGGATTCCTGAGTGGGAGTTTGTCAATACCCCTCCCTTAGTGAAATTATGGTACCAGT

    LeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgG
3388 TAGAGAAAGAACCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCTAATAGGG

     luThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValValSer 623
     AGACTAAATTAGGAAAAGCAGGATATGTTACTGACAGAGGAAGACAAAAAGTTGTCTCCA

    IleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspS
3508 TAGCTGACACAACAAATCAGAAGACTGAATTACAAGCAATTCATCTAGCTTTGCAGGATT

     erGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla 663
     CGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCAC

    GlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIleLysLysG
3628 AACCAGATAAGAGTGAATCAGAGTTAGTCAGTCAAATAATAGAGCAGTTAATAAAAAAGG

     luLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 703
     AAAAGGTCTACCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAG

    AspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIleAspLysAlaG
3748 ATAAATTAGTCAGTGCTGGAATCAGGAAAGTACTATTTTTGAATGGAATAGATAAGGCCC

     lnGluGluHisGluLysTyrHisSerAsnTrpArgAlaMetAlaSerAspPheAsnLeu 743
     AAGAAGAACATGAGAAATATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTGC

    ProProValValAlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGluA
3868 CACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAG

     laMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAspCysThrHisLeu 783
     CCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATCTAG

    GluGlyLysIleIleLeuValAlaValHisValAlaSerGlyTyrIleGluAlaGluValI
3988 AAGGAAAAATTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTTA

     leProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLeuLysLeuAlaGlyArgTrp 823
     TTCCAGCAGAGACAGGGCAGGAAACAGCATATTTTCTCTTAAAATTAGCAGGAAGATGGC

    ProValLysThrIleHisThrAspAsnGlySerAsnPheThrSerThrThrValLysAlaA
4108 CAGTAAAAACAATACATACAGACAATGGCAGCAATTTCACCAGTACTACGGTTAAGGCCG

     laCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGln 863
     CCTGTTGGTGGGCAGGGATCAAGCAGGAATTTGGCATTCCCTACAATCCCCAAAGTCAAG

    GlyValValGluSerMetAsnAsnGluLeuLysLysIleIleGlyGlnValArgAspGlnA
4228 GAGTAGTAGAATCTATGAATAATGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAGG

     laGluHisLeuLysThrAlaValGlnMetAlaValPheIleHisAsnPheLysArgLys 903
     CTGAACACCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAG

    GlyGlyIleGlyGlyTyrSerAlaGlyGluArgIleValAspIleIleAlaThrAspIleG
4348 GGGGGATTGGGGGATACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATAC

     lnThrLysGluLeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArg 943
     AAACTAAAGAACTACAAAAGCAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGG
```

FIG. 5D

```
    AspAsnLysAspProLeuTrpLysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAlaV
4468 ACAACAAAGATCCCCTTTGGAAAGGACCAGCAAAGCTTCTCTGGAAAGGTGAAGGGGCAG

     alValIleGlnAspAsnSerAspIleLysValValProArgArgLysAlaLysIleIle 983
     TAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAAGCAAAAATCATTA

    ArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArgGlnAspGluAspA
4588 GGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATT

     M
     AGAACATGGAAAAGTTTAGTAAAACACCATATGTATATTTCAAAGAAAGCTAAAGGATGG
4708 TTTTATAGACATCACTATGAAAGTACTCATCCAAGAGTAAGTTCAGAAGTACACATCCCC
     CTAGGGGATGCTAAATTGGTAATAACAACATATTGGGGTCTGCATACAGGAGAAAGAGAA
4828 TGGCATTTGGGCCAGGGAGTCGCCATAGAATGGAGGAAAAAGAAATATAGCACACAAGTA
     GACCCTGGCCTAGCAGACCAACTAATTCATCTGCATTATTTTGATTGTTTTTCAGAATCT
4948 GCTATAAAAAATGCCATATTAGGATATAGAGTTAGTCCTAGGTGTGAATATCAAGCAGGA
     CATAACAAGGTAGGATCTCTACAATACTTGGCACTAGCAGCATTAATAACACCAAAAAAG
5068 ACAAAGCCACCTTTGCCTAGTGTTAAGAAACTGACAGAGGATAGATGGAACAAGCCCCAG
     AAGACCAAGGGCCACAGAGGGAGCCATACAATGAATGGACACTAGAGCTTTTAGAGGAGC
5188 TTAAGAGAGAAGCTGTTAGACATTTTCCTAGGCCATGGCTCCATAGCTTAGGACAATATA
     TCTATGAAACTTATGGGGATACTTGGGCAGGAGTGGAAGCCATAATAAGAATTCTGCAAC
5308 AACTGCTGTTTATTCATTTCAGAATTGGGTGTCAACATAGCAGAATAGGCATTATTCAAC
     AGAGGAGAGCAAGAAGAAATGGAGCCAGTAGATCCTAATCTAGAGCCCTGGAAGCATCCA
5428 GGAAGTCAGCCTAGGACTGCTTGTAACAATTGCTATTGTAAAAAGTGTTGCTTTCATTGC
     TACGCGTGTTTCACAAGAAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAG
5548 CGACGAAGAGCTCCTCAGGACAGTCAGACTCATCAAGCTTCTCTATCAAAGCAGTAAGTA
     GTAAATGTAATGCAATCTTTACAAATATTAGCAATAGTATCATTAGTAGTAGTAGCAATA

5668 ATAGCAATAGTTGTGTGGACCATAGTACTCATAGAATATAGGAAAATATTAAGACAAAGA

                                                        MetLysVal   3
     AAATAGACAGATTAATTGATAGAATAAGAGAAAAAGCAGAAGACAGTGGCAATGAAAGTG

     LysGlyThrArgArgAsnTyrGlnHisLeuTrpArgTrpGlyThrLeuLeuLeuGlyMet
5788 AAGGGGACCAGGAGGAATTATCAGCACTTGTGGAGATGGGGCACCTTGCTCCTTGGGATG

     LeuMetIleCysSerAlaThrGluLysLeuTrpValThrValTyrTyrGlyValProVal  43
     TTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTTTATTATGGAGTACCTGTG

     TrpLysGluAlaThrThrThrLeuPheCysAlaSerAspAlaArgAlaTyrAspThrGlu
5908 TGGAAAGAAGCAACTACCACTCTATTTTGTGCATCAGATGCTAGAGCATATGATACAGAG

     ValHisAsnValTrpAlaThrHisAlaCysValProThrAspProAsnProGlnGluVal  83
     GTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTA
```

FIG. 5E

```
     ValLeuGlyAsnValThrGluAsnPheAsnMetTrpLysAsnAsnMetValGluGlnMet
6028 GTATTGGGAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACATGGTAGAACAGATG

     GlnGluAspIleIleSerLeuTrpAspGlnSerLeuLysProCysValLysLeuThrPro 123
     CAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCA

     LeuCysValThrLeuAsnCysThrAspLeuGlyLysAlaThrAsnThrAsnSerSerAsn
6148 CTCTGTGTTACTTTAAATTGCACTGATTTGGGGAAGGCTACTAATACCAATAGTAGTAAT

     TrpLysGluGluIleLysGlyGluIleLysAsnCysSerPheAsnIleThrThrSerIle 163
     TGGAAAGAAGAAATAAAAGGAGAAATAAAAAACTGCTCTTTCAATATCACCACAAGCATA

     ArgAspLysIleGlnLysGluAsnAlaLeuPheArgAsnLeuAspValValProIleAsp
6268 AGAGATAAGATTCAGAAAGAAAATGCACTTTTTCGTAACCTTGATGTAGTACCAATAGAT

     AsnAlaSerThrThrThrAsnTyrThrAsnTyrArgLeuIleHisCysAsnArgSerVal 203
     AATGCTAGTACTACTACCAACTATACCAACTATAGGTTGATACATTGTAACAGATCAGTC

     IleThrGlnAlaCysProLysValSerPheGluProIleProIleHisTyrCysThrPro
6388 ATTACACAGGCCTGTCCAAAGGTATCATTTGAGCCAATTCCCATACATTATTGTACCCCG

     AlaGlyPheAlaIleLeuLysCysAsnAsnLysThrPheAsnGlyLysGlyProCysThr 243  E
     GCTGGTTTTGCGATTCTAAAGTGTAATAATAAAACGTTCAATGGAAAAGGACCATGTACA

     AsnValSerThrValGlnCysThrHisGlyIleArgProIleValSerThrGlnLeuLeu
6508 AATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAATAGTGTCAACTCAACTGCTG

     LeuAsnGlySerLeuAlaGluGluGluValValIleArgSerAspAsnPheThrAsnAsn 283  N
     TTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGACAATTTCACGAACAAT

     AlaLysThrIleIleValGlnLeuAsnGluSerValAlaIleAsnCysThrArgProAsn
6628 GCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATTAACTGTACAAGACCCAAC

     AsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAlaPheHisThrThrGlyArg 323  V
     AACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCATTTCATACAACAGGAAGA

     IleIleGlyAspIleArgLysAlaHisCysAsnIleSerArgAlaGlnTrpAsnAsnThr
6748 ATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGAGCACAATGGAATAACACT

     LeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsnAsnLysThrIleValPhe 363
     TTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAATAATAAAACAATAGTCTTT

     AsnGlnSerSerGlyGlyAspProGluIleValMetHisSerPheAsnCysArgGlyGlu
6868 AATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGGGAA

     PhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrpArgLeuAsnHisThrGlu 403
     TTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGGAGGTTAAATCACACTGAA

     GlyThrLysGlyAsnAspThrIleIleLeuProCysArgIleLysGlnIleIleAsnMet
6988 GGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATAAAACAAATTATAAACATG

     TrpGlnGluValGlyLysAlaMetTyrAlaProProIleGlyGlyGlnIleSerCysSer 443
     TGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGAGGACAAATTAGTTGTTCA

     SerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThrAsnValThrAsnAspThr
7108 TCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACAAATGTAACTAATGACACC

     GluValPheArgProGlyGlyGlyAspMetArgAspAsnTrpArgSerGluLeuTyrLys 483
     GAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAA
```

FIG. 5F

```
    TyrLysValIleLysIleGluProLeuGlyIleAlaProThrLysAlaLysArgArgVal
7228 TATAAAGTAATAAAAATTGAACCATTAGGAATAGCACCCACCAAGGCAAAGAGAAGAGTG

    ValGlnArgGluLysArgAlaValGlyIleValGlyAlaMetPheLeuGlyPheLeuGly 523
    GTGCAGAGAGAAAAAAGAGCAGTGGGAATAGTAGGAGCTATGTTCCTTGGGTTCTTGGGA

    AlaAlaGlySerThrMetGlyAlaValSerLeuThrLeuThrValGlnAlaArgGlnLeu
7348 GCAGCAGGAAGCACTATGGGCGCAGTGTCATTGACGCTGACGGTACAGGCCAGACAATTA

    LeuSerGlyIleValGlnGlnGlnAsnAsnLeuLeuArgAlaIleGluAlaGlnGlnHis 563
    TTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAACAT

    LeuLeuGlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaValGlu
7468 CTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGTGGAA

    ArgTyrLeuArgAspGlnGlnLeuLeuGlyIleTrpGlyCysSerGlyLysLeuIleCys 603
    AGATACCTAAGGGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTGGAAAACTCATTTGC

    ThrThrAlaValProTrpAsnAlaSerTrpSerAsnLysSerLeuGluAspIleTrpAsp
7588 ACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAAGACATTTGGGAT

    AsnMetThrTrpMetGlnTrpGluArgGluIleAspAsnTyrThrAsnThrIleTyrThr 643
    AACATGACCTGGATGCAGTGGGAAAGAGAAATTGACAATTACACAAACACAATATACACC

    LeuLeuGluGluSerGlnAsnGlnGlnGluLysAsnGluGlnGluLeuLeuGluLeuAsp
7708 TTACTTGAAGAATCGCAGAACCAACAAGAAAAGAATGAACAAGAATTATTAGAATTGGAT

    LysTrpAlaSerLeuTrpAsnTrpPheSerIleThrAsnTrpLeuTrpTyrIleLysIle 683
    AAGTGGGCAAGTTTGTGGAATTGGTTTAGCATAACAAACTGGCTGTGGTATATAAAGATA

    PheIleMetIleValGlyGlyLeuValGlyLeuArgIleValPheAlaValLeuSerIle      E
7828 TTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTGCTTTCTATA

    ValAsnArgValArgGlnGlyTyrSerProLeuSerPheGlnThrArgLeuProValPro 723
    GTGAATAGAGTTAGGCAGGGATACTCACCATTGTCATTTCAGACCCGCCTCCCAGTCCCG      N

    ArgGlyProAspArgProAspGlyIleGluGluGluGlyGlyGluArgAspArgAspArg
7948 AGGGGACCCGACAGGCCCGACGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGA

    SerValArgLeuValAspGlyPheLeuAlaLeuIleTrpGluAspLeuArgSerLeuCys 763  V
    TCCGTTCGATTAGTGGATGGATTCTTAGCACTTATCTGGGAAGATCTGCGGAGCCTGTGC

    LeuPheSerTyrArgArgLeuArgAspLeuLeuLeuIleAlaAlaArgThrValGluIle
8068 CTCTTCAGCTACCGCCGCTTGAGAGACTTACTCTTGATTGCAGCGAGGACTGTGGAAATT

    LeuGlyHisArgGlyTrpGluAlaLeuLysTyrTrpTrpSerLeuLeuGlnTyrTrpIle 803
    CTGGGGCACAGGGGGTGGGAAGCCCTCAAATATTGGTGGAGTCTCCTGCAGTATTGGATT

    GlnGluLeuLysAsnSerAlaValSerTrpLeuAsnAlaThrAlaIleAlaValThrGlu
8188 CAGGAACTAAAGAATAGTGCTGTTAGCTGGCTCAACGCCACAGCTATAGCAGTAACTGAG

    GlyThrAspArgValIleGluValAlaGlnArgAlaTyrArgAlaIleLeuHisIleHis 843
    GGGACAGATAGGGTTATAGAAGTAGCACAAAGAGCTTATAGAGCTATTCTCCACATACAT

    ArgArgIleArgGlnGlyLeuGluArgLeuLeuLeuOC
8308 AGAAGAATTAGACAGGGCTTGGAAAGGCTTTTGCTATAAGATGGGTGGCAAGTGGTCAAA

    ACGTAGTATGGGTGGATGGTCTGCTATAAGGGAAAGAATGAGACGAGCTGAGCCACGAGC
```

FIG. 5G

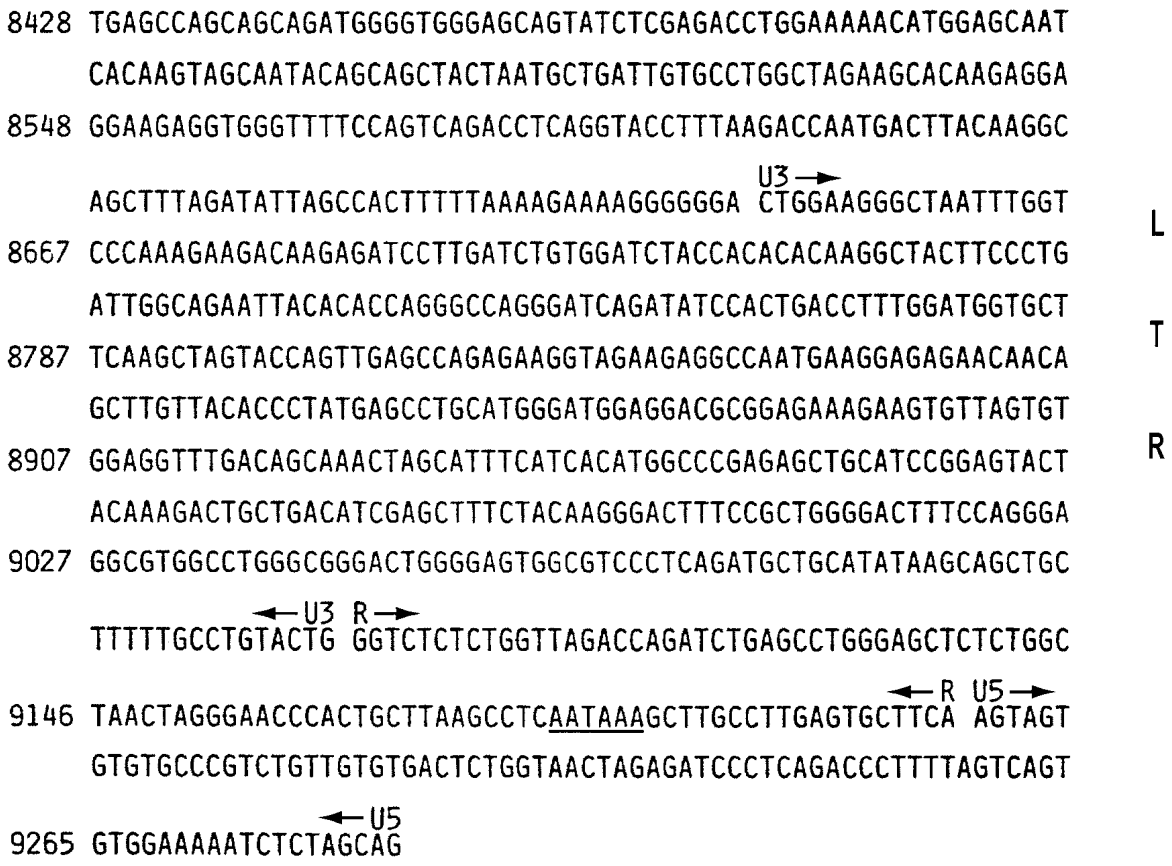

8428 TGAGCCAGCAGCAGATGGGGTGGGAGCAGTATCTCGAGACCTGGAAAAACATGGAGCAAT
CACAAGTAGCAATACAGCAGCTACTAATGCTGATTGTGCCTGGCTAGAAGCACAAGAGGA
8548 GGAAGAGGTGGGTTTTCCAGTCAGACCTCAGGTACCTTTAAGACCAATGACTTACAAGGC
U3
AGCTTTAGATATTAGCCACTTTTTAAAAGAAAAGGGGGGA CTGGAAGGGCTAATTTGGT
8667 CCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTG
ATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCT
8787 TCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCCAATGAAGGAGAGAACAACA
GCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCGGAGAAAGAAGTGTTAGTGT
8907 GGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGAGTACT
ACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGA
9027 GGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATGCTGCATATAAGCAGCTGC
U3 R
TTTTTGCCTGTACTG GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGC
R U5
9146 TAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCA AGTAGT
GTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGT
U5
9265 GTGGAAAAATCTCTAGCAG
L
T
R

FIG. 5H

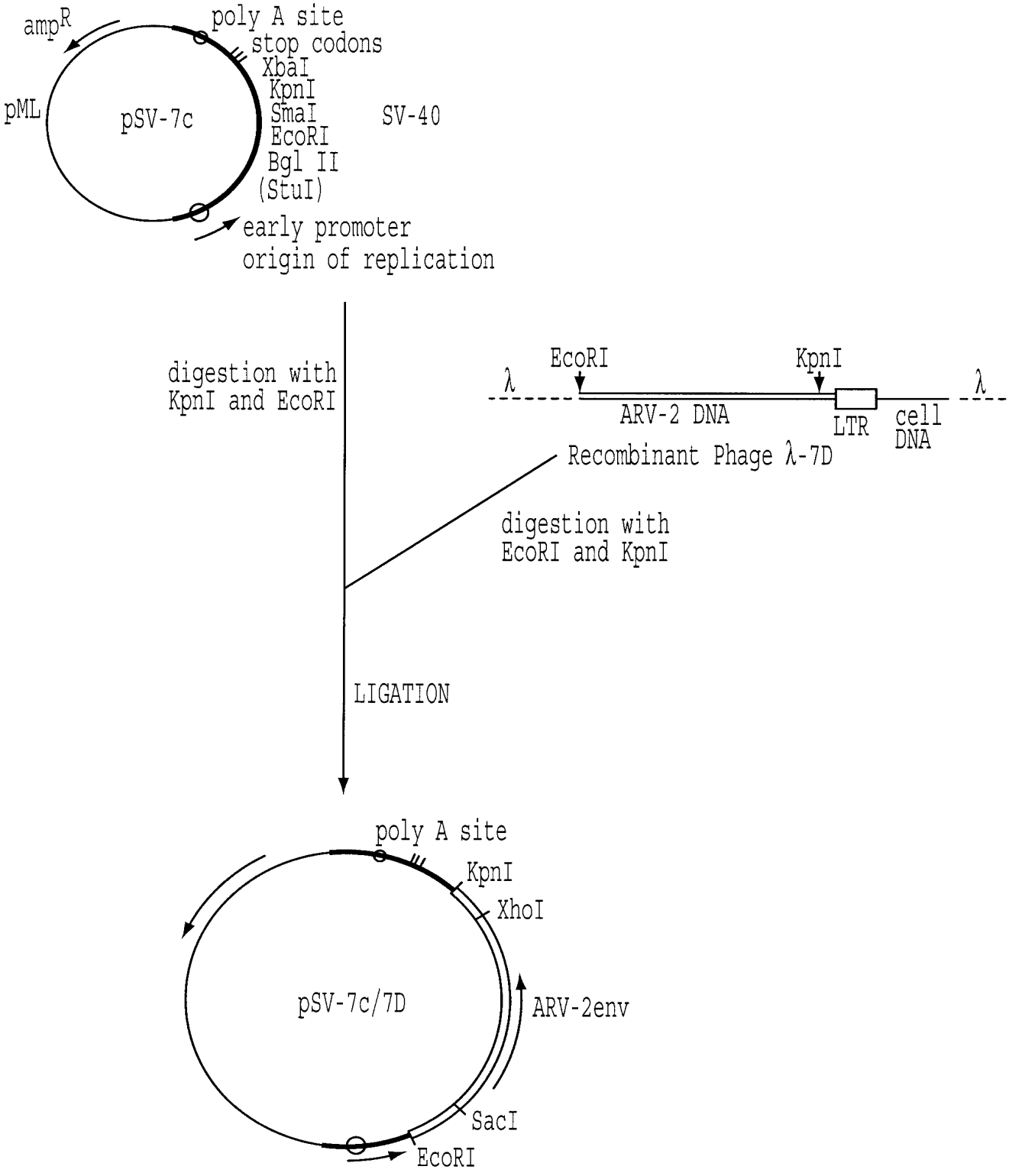

amp^R
poly A site
stop codons
XbaI
KpnI
SmaI
EcoRI
Bgl II
(StuI)
pML
pSV-7c
SV-40
early promoter
origin of replication
digestion with
KpnI and EcoRI
EcoRI
KpnI
λ
λ
ARV-2 DNA
LTR
cell
DNA
Recombinant Phage λ-7D
digestion with
EcoRI and KpnI
LIGATION
poly A site
KpnI
XhoI
pSV-7c/7D
ARV-2env
SacI
EcoRI

FIG. 6

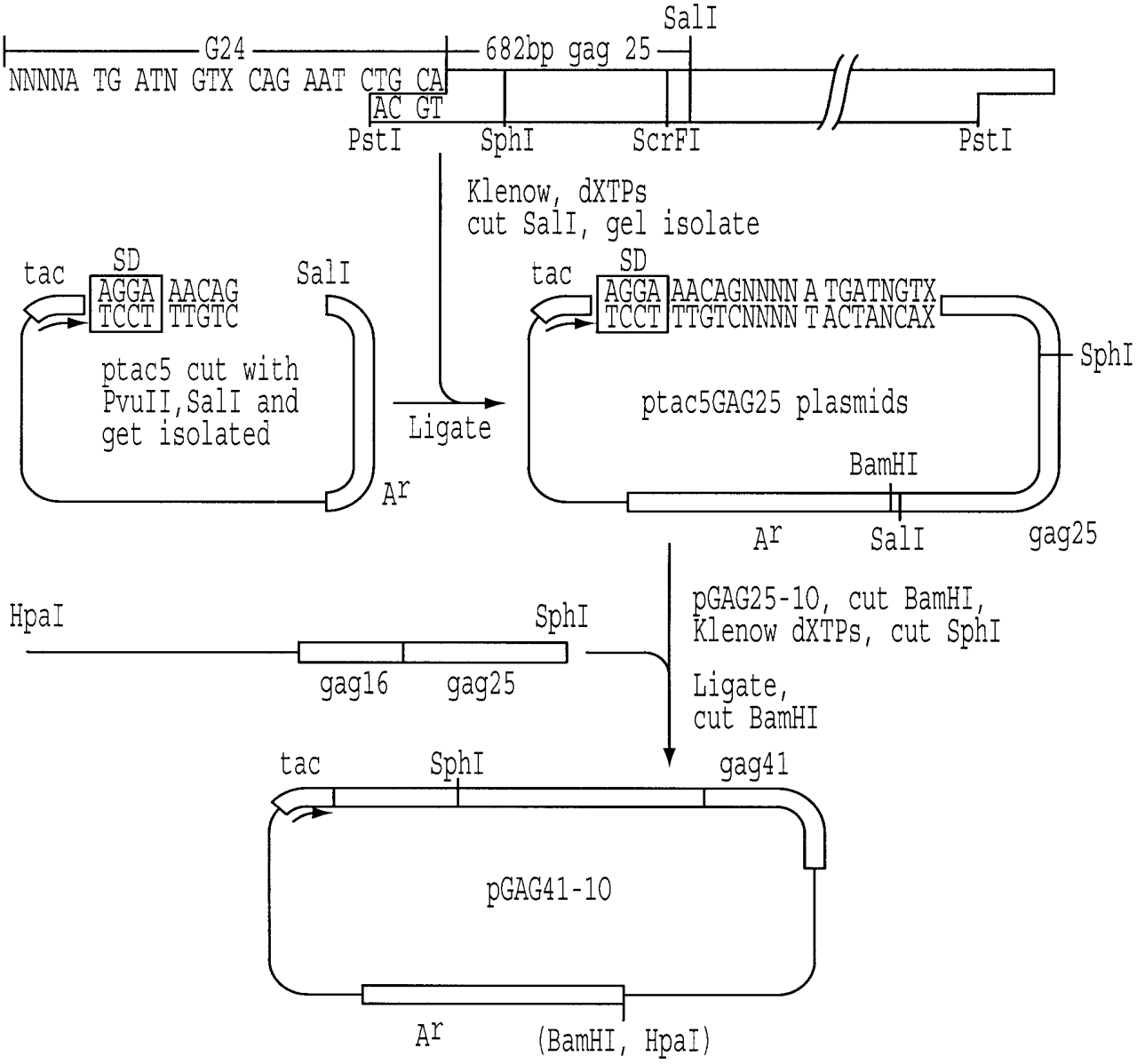

SalI
G24
682bp gag 25
NNNNA TG ATN GTX CAG AAT CTG CA
AC GT
PstI
SphI
ScrFI
PstI
Klenow, dXTPs
cut SalI, gel isolate
tac
SD
AGGA
TCCT
AACAG
TTGTC
SalI
ptac5 cut with
PvuII,SalI and
get isolated
Ar
Ligate
tac
SD
AGGA
TCCT
AACAGNNNN A TGATNGTX
TTGTCNNNN T ACTANCAX
SphI
ptac5GAG25 plasmids
BamHI
Ar
SalI
gag25
pGAG25-10, cut BamHI,
Klenow dXTPs, cut SphI
HpaI
SphI
gag16
gag25
Ligate,
cut BamHI
tac
SphI
gag41
pGAG41-10
Ar
(BamHI, HpaI)

FIG. 7

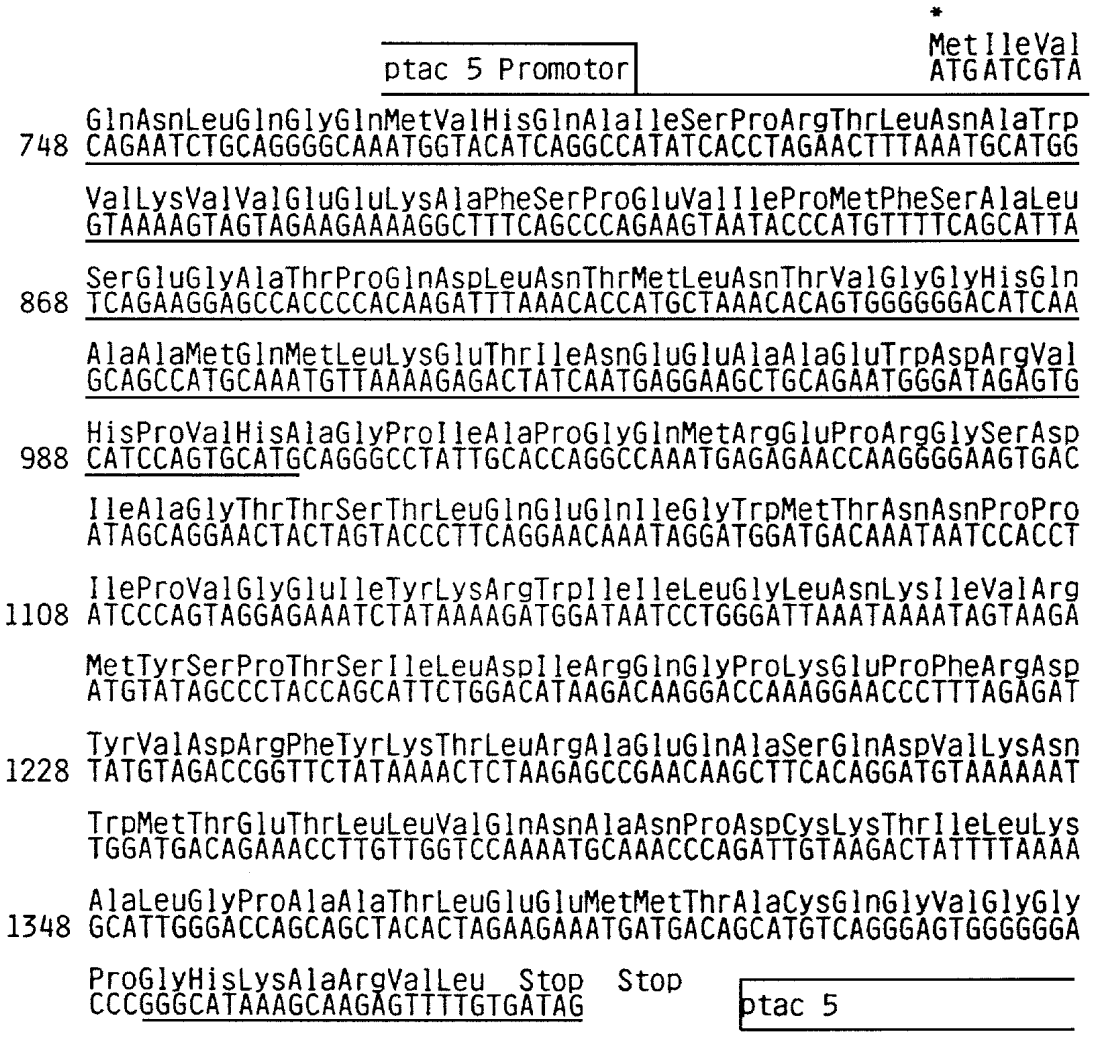

*
MetIleVal
ATGATCGTA
ptac 5 Promotor
GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrp
748 CAGAATCTGCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG
ValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu 181
GTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA
SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGln
868 TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAA
AlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal 221
GCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG
HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAsp
988 CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGAC
IleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProPro 261
ATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT
IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArg
1108 ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGA
MetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp 301
ATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT
TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsn
1228 TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAAT
TrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys 341
TGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA
AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGly
1348 GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGA
ProGlyHisLysAlaArgValLeu Stop Stop
CCCGGGCATAAAGCAAGAGTTTTGTGATAG
ptac 5

FIG. 8

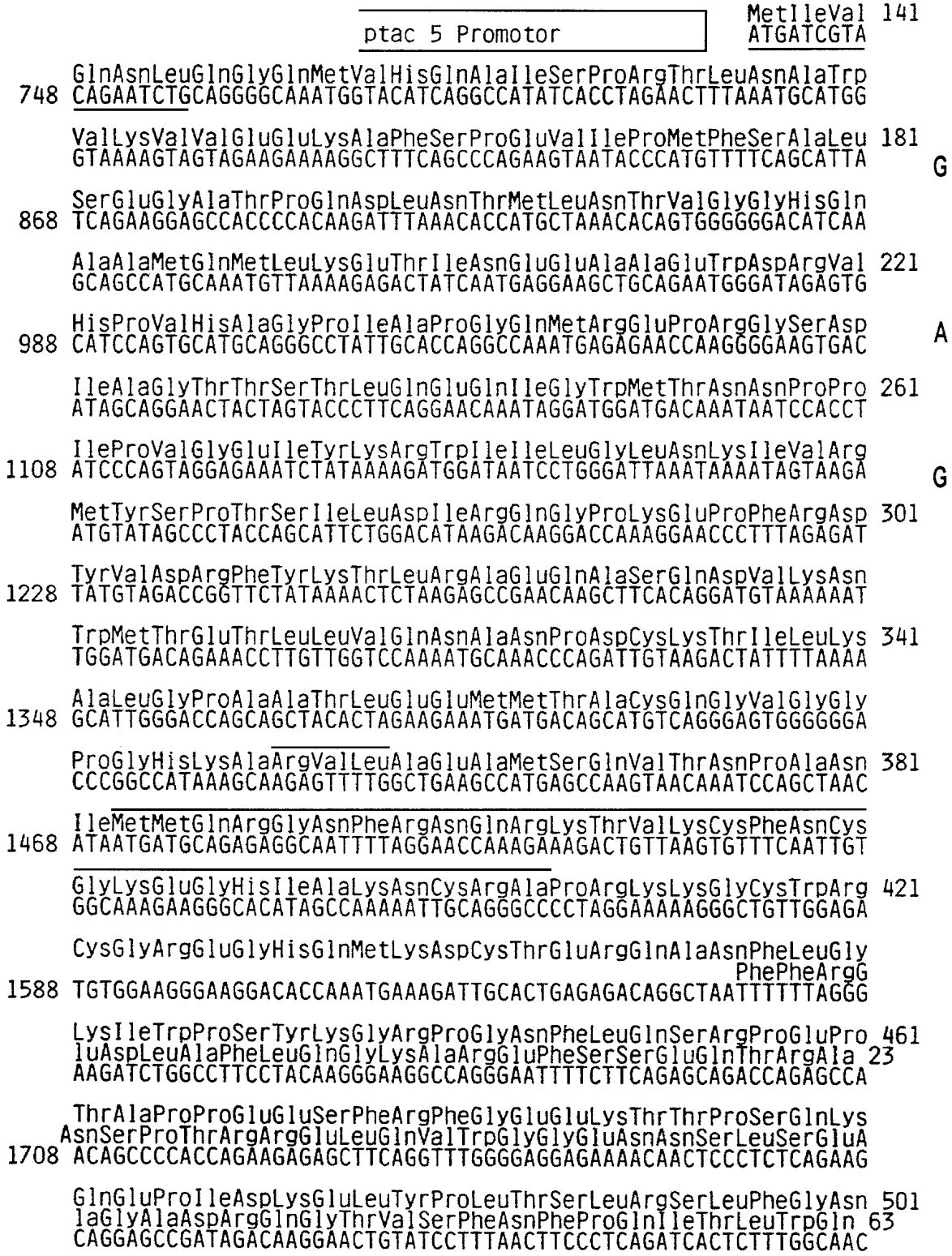

ptac 5 Promotor
MetIleVal 141
ATGATCGTA
GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrp
748 CAGAATCTGCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG
ValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu 181
GTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA G
SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGln
868 TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAA
AlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal 221
GCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG
HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAsp
988 CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGAC A
IleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProPro 261
ATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT
IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArg
1108 ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGA G
MetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp 301
ATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT
TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsn
1228 TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAAT
TrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys 341
TGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA
AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGly
1348 GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGA
ProGlyHisLysAlaArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaAsn 381
CCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCCATGAGCCAAGTAACAAATCCAGCTAAC
IleMetMetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCys
1468 ATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGACTGTTAAGTGTTTCAATTGT
GlyLysGluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpArg 421
GGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAGA
CysGlyArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGly
PhePheArgG
1588 TGTGGAAGGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGG
LysIleTrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluPro 461
luAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAla 23
AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA
ThrAlaProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLys
AsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSerLeuSerGluA
1708 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG
GlnGluProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsn 501
laGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGln 63
CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC

FIG. 9A

```
     AspProSerSerGlnOC
    ArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyA
1828 GACCCCTCGTCACAATAAGGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAG

     laAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLysMetIle 103
     CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG

    GlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProValGluIleCysG
1948 GGGGAATTGGAGGTTTTATCAAAGTAAGACAGTACGATCAGATACCTGTAGAAATCTGTG

     lyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArg 143
     GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA

    AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValP
2068 ATCTGTTGACTCAGATTGGTTGTACTTTAAATTTCCCCATTAGTCCTATTGAAACTGTAC

     roValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 183
     CAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAGCAATGGCCATTGACAGAAG

    GluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerL
2188 AAAAAATAAAAGCATTAGTAGAGATATGTACAGAAATGGAAAAGGAAGGGAAAATTTCAA

     ysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 223
     AAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCTATAAAGAAAAAAGACAGTA

    ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpG
2308 CTAAATGGAGAAAACTAGTAGATTTCAGAGAACTTAATAAAAGAACTCAAGACTTCTGGG

     luValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVal 263
     AAGTTCAGTTAGGAATACCACACCCCGCAGGGTTAAAAAAGAAAAAATCAGTAACAGTAT

    LeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLysTyrThrA
2428 TGGATGTGGGTGATGCATACTTTTCAGTTCCCTTAGATAAAGACTTTAGAAAGTATACTG

     laPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 303
     CATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGC

    LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuG
2548 TGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAG

     luProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyr 343
     AGCCTTTTAGAAAACAGAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATG

    ValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisL
2668 TAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAACTGAGACAGCATC

     euLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 383
     TGTTGAGGTGGGGATTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTT

    TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMetLeuProGluL
2788 GGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAATGCTGCCAGAAA

     ysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 423
     AAGACAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGAAAATTGAATTGGGCAAGTC

    GlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThrLysAlaL
2908 AGATTTATGCAGGGATTAAAGTAAAGCAGTTATGTAAACTCCTTAGAGGAACCAAAGCAC

     euThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 463 P
     TAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGGGAGA   O
                                                                    L
```

FIG. 9B

```
    IleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeuValAlaGluI
3028 TTCTAAAAGAACCAGTACATGAAGTATATTATGACCCATCAAAAGACTTAGTAGCAGAAA

     leGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 503
     TACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAATC

    LeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuT
3148 TGAAAACAGGAAAGTATGCAAGGATGAGGGGTGCCCACACTAATGATGTAAAACAGTT

     hrGluAlaValGluLysValSerThrGluSerIleValIleTrpGlyLysIleProLys 543
     [ptac 5]
```

FIG. 9C

ARV GAG p16 - synthetic Parts A and B

```
5'      arv 234          3'
    MetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLys
TATTATGCAAAGAGGTAACTTCAGGAATCAAAGAAAGACCGTTAAGTGTTTCAACTGTGGTAAG
ATAATACGTTTCTCCATTGAAGTCCTTAGTTTCTTTCTGGCAATTCACAAAGTTGACACCATTC
3'          ^ arv 235     ^ 5'
    10 mnl1, 23 hinf1,

    GluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysAlaCysTrpArgCysGly
 63 GAAGGTCACATCGCTAAGAACTGTAGAGCTCCAAGAAAGAAGGCTTGTTGGAGATGTGGT
    CTTCCAGTGTAGCGATTCTTGACATCTCGAGGTTCTTTCTTCCGAACAACCTCTACACCA

    76 dde1, 88 ban2 hgiA hgiJ11 sac1 sduI, 89 alu1,

    ArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIle
123 AGAGAAGGTCACCAAATGAAGGACTGTACCGAAAGACAAGCTAACTTCTTGGGTAAGATC
    TCTCTTCCAGTGGTTTACTTCCTGACATGGCTTTCTGTTCGATTGAAGAACCCATTCTAG

    129 bstE2, 131 hph, 148 rsaI, 161 alu1, 178 bgl11 xho2, 179
    sau3a,

    TrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAla
183 TGGCCATCTTACAAGGGTAGACCAGGTAACTTCTTGCAATCCAGACCAGAACCAACCGCT
    ACCGGTAGAATGTTCCCATCTGGTCCATTGAAGAACGTTAGGTCTGGTCTTGGTTGGCGA

    183 ball cfr1 hae1, 184 hae111, 199 acc1, 204 apy1 ecor11 sc
    rF1,

    ProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLysGlnGlu
243 CCACCTGAAGAAAGTTTCAGGTTCGGTGAAGAAAAGACCACCCCATCTCAAAAGCAAGAA
    GGTGGACTTCTTTCAAAGTCCAAGCCACTTCTTTTCTGGTGGGGTAGAGTTTTCGTTCTT

     249 mbo11, 267 hph, 270 mbo11,

    ProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsnAspPro
303 CCAATCGACAAGGAATTGTACCCATTGACCTCTTTGAGATCCTTGTTCGGTAACGATCCC
    GGTTAGCTGTTCCTTAACATGGGTAACTGGAGAAACTCTAGGAACAAGCCATTGCTAGGG

    307 taq1, 320 rsaI, 331 mnl1, 339 xho2, 340 sau3a, 357 sau3a,
    361 mnl1, 362 aval xho1,

    SerSerGlnOP AM
363 TCGAGCCAATGATAG
    AGCTCGGTTACTATCAGCT

    363 taq1, 377 acc1 hindl1 sal1
```

FIG. 10

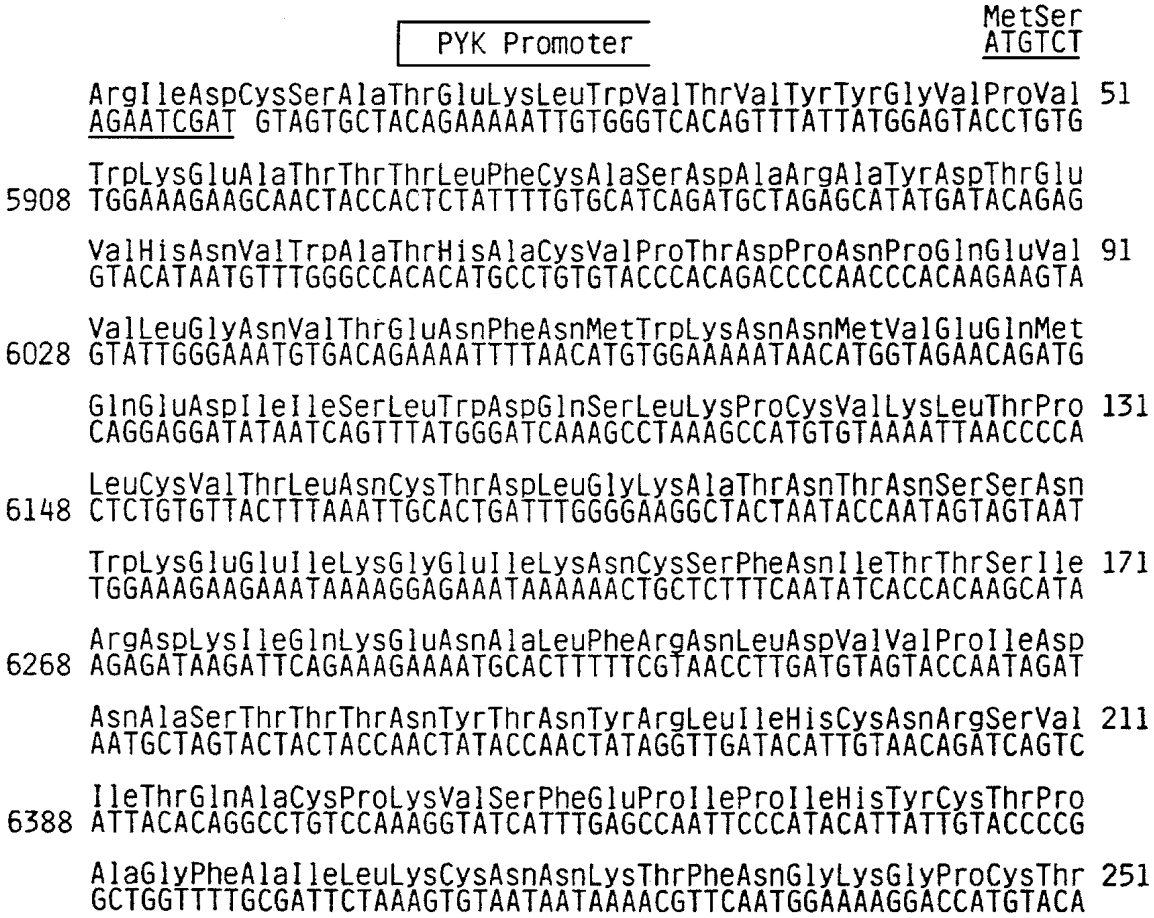

PYK Promoter
MetSer
ATGTCT
ArgIleAspCysSerAlaThrGluLysLeuTrpValThrValTyrTyrGlyValProVal 51
AGAATCGAT GTAGTGCTACAGAAAAATTGTGGGTCACAGTTTATTATGGAGTACCTGTG
TrpLysGluAlaThrThrThrLeuPheCysAlaSerAspAlaArgAlaTyrAspThrGlu
5908 TGGAAAGAAGCAACTACCACTCTATTTTGTGCATCAGATGCTAGAGCATATGATACAGAG
ValHisAsnValTrpAlaThrHisAlaCysValProThrAspProAsnProGlnGluVal 91
GTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTA
ValLeuGlyAsnValThrGluAsnPheAsnMetTrpLysAsnAsnMetValGluGlnMet
6028 GTATTGGGAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACATGGTAGAACAGATG
GlnGluAspIleIleSerLeuTrpAspGlnSerLeuLysProCysValLysLeuThrPro 131
CAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCA
LeuCysValThrLeuAsnCysThrAspLeuGlyLysAlaThrAsnThrAsnSerSerAsn
6148 CTCTGTGTTACTTTAAATTGCACTGATTTGGGGAAGGCTACTAATACCAATAGTAGTAAT
TrpLysGluGluIleLysGlyGluIleLysAsnCysSerPheAsnIleThrThrSerIle 171
TGGAAAGAAGAAATAAAAGGAGAAATAAAAAACTGCTCTTTCAATATCACCACAAGCATA
ArgAspLysIleGlnLysGluAsnAlaLeuPheArgAsnLeuAspValValProIleAsp
6268 AGAGATAAGATTCAGAAAGAAAATGCACTTTTTCGTAACCTTGATGTAGTACCAATAGAT
AsnAlaSerThrThrThrAsnTyrThrAsnTyrArgLeuIleHisCysAsnArgSerVal 211
AATGCTAGTACTACTACCAACTATACCAACTATAGGTTGATACATTGTAACAGATCAGTC
IleThrGlnAlaCysProLysValSerPheGluProIleProIleHisTyrCysThrPro
6388 ATTACACAGGCCTGTCCAAAGGTATCATTTGAGCCAATTCCCATACATTATTGTACCCCG
AlaGlyPheAlaIleLeuLysCysAsnAsnLysThrPheAsnGlyLysGlyProCysThr 251
GCTGGTTTTGCGATTCTAAAGTGTAATAATAAAACGTTCAATGGAAAAGGACCATGTACA

FIG. 11A

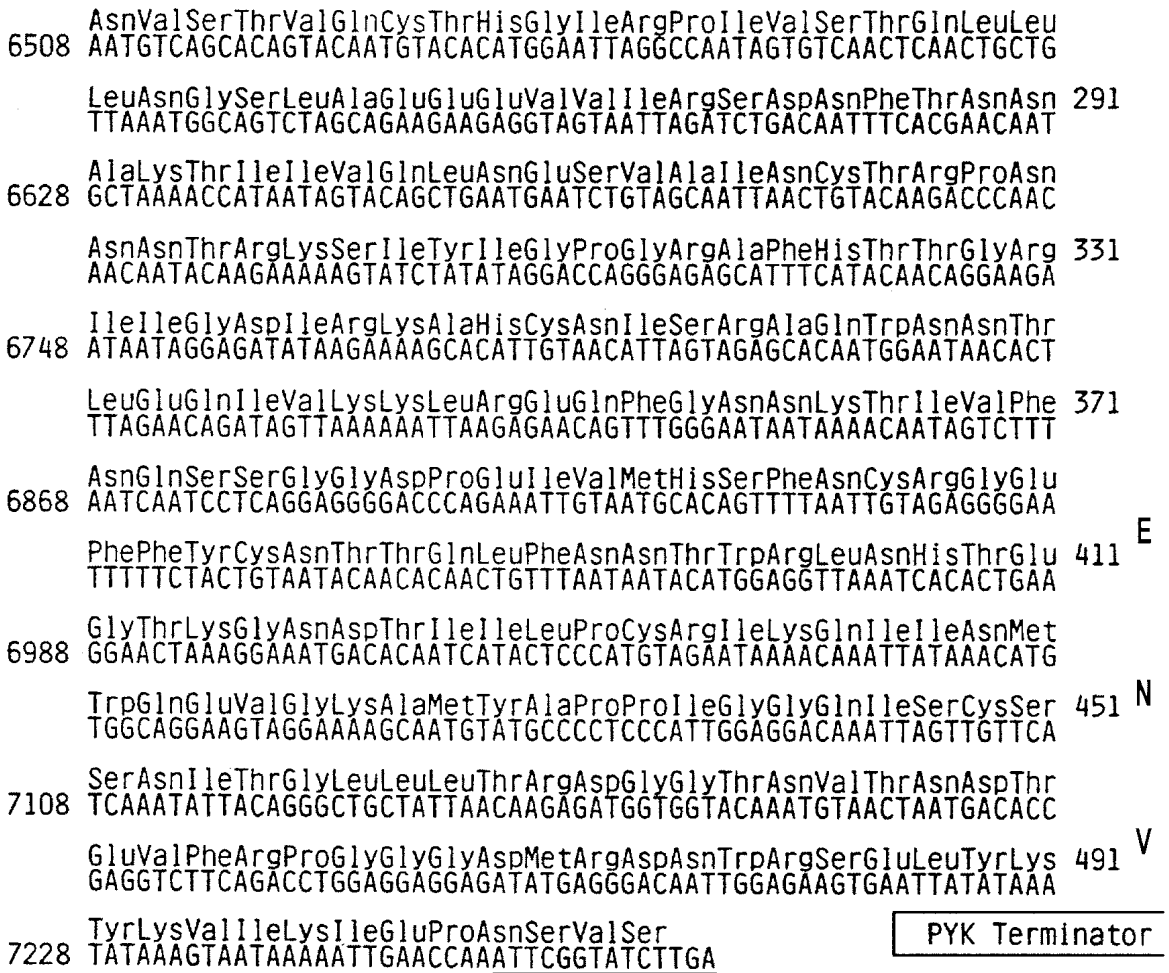

AsnValSerThrValGlnCysThrHisGlyIleArgProIleValSerThrGlnLeuLeu
6508 AATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAATAGTGTCAACTCAACTGCTG
LeuAsnGlySerLeuAlaGluGluGluValValIleArgSerAspAsnPheThrAsnAsn 291
TTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGACAATTTCACGAACAAT
AlaLysThrIleIleValGlnLeuAsnGluSerValAlaIleAsnCysThrArgProAsn
6628 GCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATTAACTGTACAAGACCCAAC
AsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAlaPheHisThrThrGlyArg 331
AACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCATTTCATACAACAGGAAGA
IleIleGlyAspIleArgLysAlaHisCysAsnIleSerArgAlaGlnTrpAsnAsnThr
6748 ATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGAGCACAATGGAATAACACT
LeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsnAsnLysThrIleValPhe 371
TTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAATAATAAAACAATAGTCTTT
AsnGlnSerSerGlyGlyAspProGluIleValMetHisSerPheAsnCysArgGlyGlu
6868 AATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGGGAA
E
PhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrpArgLeuAsnHisThrGlu 411
TTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGGAGGTTAAATCACACTGAA
GlyThrLysGlyAsnAspThrIleIleLeuProCysArgIleLysGlnIleIleAsnMet
6988 GGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATAAAACAAATTATAAACATG
TrpGlnGluValGlyLysAlaMetTyrAlaProProIleGlyGlyGlnIleSerCysSer 451
N
TGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGAGGACAAATTAGTTGTTCA
SerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThrAsnValThrAsnAspThr
7108 TCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACAAATGTAACTAATGACACC
GluValPheArgProGlyGlyGlyAspMetArgAspAsnTrpArgSerGluLeuTyrLys 491
V
GAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAA
TyrLysValIleLysIleGluProAsnSerValSer
PYK Terminator
7228 TATAAAGTAATAAAAATTGAACCAAATTCGGTATCTTGA

FIG. 11B

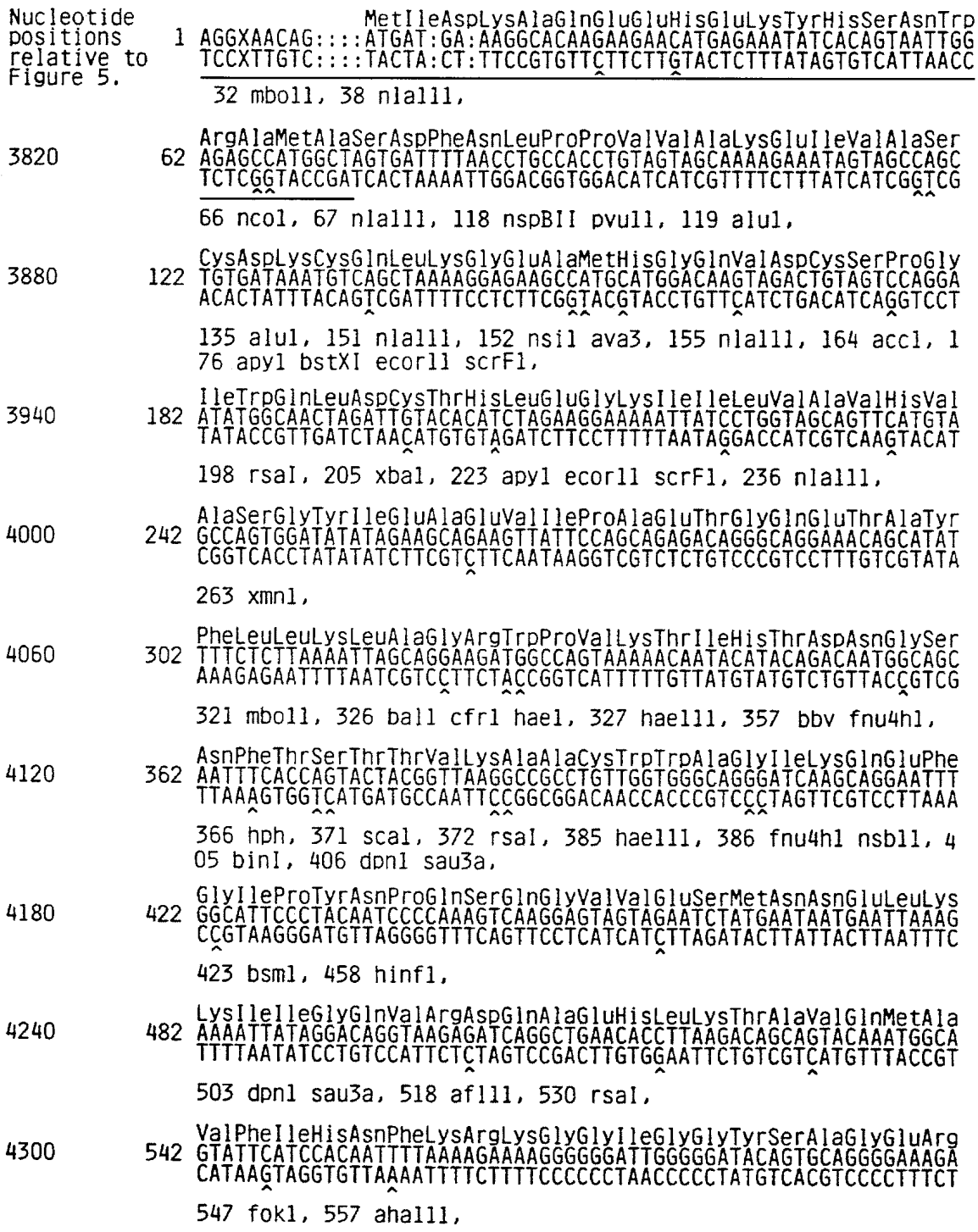

Nucleotide positions relative to Figure 5.
1 MetIleAspLysAlaGlnGluGluHisGluLysTyrHisSerAsnTrp
AGGXAACAG::::ATGAT:GA:AAGGCACAAGAAGAAGAACATGAGAAATATCACAGTAATTGG
TCCXTTGTC::::TACTA:CT:TTCCGTGTTCTTCTTGTACTCTTTATAGTGTCATTAACC
32 mboII, 38 nlaIII,
3820 62 ArgAlaMetAlaSerAspPheAsnLeuProProValValAlaLysGluIleValAlaSer
AGAGCCATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGC
TCTCGGTACCGATCACTAAAATTGGACGGTGGACATCATCGTTTTCTTTATCATCGGTCG
66 ncoI, 67 nlaIII, 118 nspBII pvuII, 119 aluI,
3880 122 CysAspLysCysGlnLeuLysGlyGluAlaMetHisGlyGlnValAspCysSerProGly
TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGA
ACACTATTTACAGTCGATTTTCCTCTTCGGTACGTACCTGTTCATCTGACATCAGGTCCT
135 aluI, 151 nlaIII, 152 nsiI ava3, 155 nlaIII, 164 accI, 176 apyI bstXI ecorII scrFI,
3940 182 IleTrpGlnLeuAspCysThrHisLeuGluGlyLysIleIleLeuValAlaValHisVal
ATATGGCAACTAGATTGTACACATCTAGAAGGAAAAATTATCCTGGTAGCAGTTCATGTA
TATACCGTTGATCTAACATGTGTAGATCTTCCTTTTTAATAGGACCATCGTCAAGTACAT
198 rsaI, 205 xbaI, 223 apyI ecorII scrFI, 236 nlaIII,
4000 242 AlaSerGlyTyrIleGluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyr
GCCAGTGGATATATAGAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAAACAGCATAT
CGGTCACCTATATATCTTCGTCTTCAATAAGGTCGTCTCTGTCCCGTCCTTTGTCGTATA
263 xmnI,
4060 302 PheLeuLeuLysLeuAlaGlyArgTrpProValLysThrIleHisThrAspAsnGlySer
TTTCTCTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCAGC
AAAGAGAATTTTAATCGTCCTTCTACCGGTCATTTTTGTTATGTATGTCTGTTACCGTCG
321 mboII, 326 ball cfrI haeI, 327 haeIII, 357 bbv fnu4hI,
4120 362 AsnPheThrSerThrThrValLysAlaAlaCysTrpTrpAlaGlyIleLysGlnGluPhe
AATTTCACCAGTACTACGGTTAAGGCCGCCTGTTGGTGGGCAGGGATCAAGCAGGAATTT
TTAAAGTGGTCATGATGCCAATTCCGGCGGACAACCACCCGTCCCTAGTTCGTCCTTAAA
366 hph, 371 scaI, 372 rsaI, 385 haeIII, 386 fnu4hI nsbII, 405 binI, 406 dpnI sau3a,
4180 422 GlyIleProTyrAsnProGlnSerGlnGlyValValGluSerMetAsnAsnGluLeuLys
GGCATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAATGAATTAAAG
CCGTAAGGGATGTTAGGGGTTTCAGTTCCTCATCATCTTAGATACTTATTACTTAATTTC
423 bsmI, 458 hinfI,
4240 482 LysIleIleGlyGlnValArgAspGlnAlaGluHisLeuLysThrAlaValGlnMetAla
AAAATTATAGGACAGGTAAGAGATCAGGCTGAACACCTTAAGACAGCAGTACAAATGGCA
TTTTAATATCCTGTCCATTCTCTAGTCCGACTTGTGGAATTCTGTCGTCATGTTTACCGT
503 dpnI sau3a, 518 afIII, 530 rsaI,
4300 542 ValPheIleHisAsnPheLysArgLysGlyGlyIleGlyGlyTyrSerAlaGlyGluArg
GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGATACAGTGCAGGGGAAAGA
CATAAGTAGGTGTTAAAATTTTCTTTTCCCCCCTAACCCCCTATGTCACGTCCCCTTTCT
547 fokI, 557 ahaIII,

FIG. 12A

```
                IleValAspIleIleAlaThrAspIleGlnThrLysGluLeuGlnLysGlnIleThrLys
4360        602 ATAGTAGACATAATAGCAACAGACATACAAACTAAAGAACTACAAAAGCAAATTACAAAA
                TATCATCTGTATTATCGTTGTCTGTATGTTTGATTTCTTGATGTTTTCGTTTAATGTTTT

                605 accl,

                IleGlnAsnPheArgValTyrTyrArgAspAsnLysAspProLeuTrpLysGlyProAla
4420        662 ATTCAAAATTTTCGGGTTTATTACAGGGACAACAAAGATCCCCTTTGGAAAGGACCAGCA
                TAAGTTTTAAAAGCCCAAATAATGTCCCTGTTGTTTCTAGGGGAAACCTTTCCTGGTCGT

                697 xho2, 698 dpnl sau3a, 713 asul ava2,

                LysLeuLeuTrpLysGlyGluGlyAlaValValIleGlnAspAsnSerAspIleLysVal
4480        722 AAGCTTCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTA
                TTCGAAGAGACCTTTCCACTTCCCCGTCATCATTATGTTCTATTATCACTGTATTTTCAT

                722 hindlll, 723 alul, 737 hph,

                ValProArgArgLysAlaLysIleIleArgAspTyrGlyLysGlnMetAlaGlyAspAsp
4540        782 GTGCCAAGAAGAAAAGCAAAAATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGAT
                CACGGTTCTTCTTTTCGTTTTTAGTAATCCCTAATACCTTTTGTCTACCGTCCACTACTA

                789 mboll, 833 hph,

                CysValAlaSerArgGlnAspGluAspAM
4600        842 TGTGTGGCAAGTAGACAGGATGAGGATTAGTCGACGGAATTCTTTAGTAAAACACC
                ACACACCGTTCATCTGTCCTACTCCTAATCAGCTGCCTTAAGAAATCATTTTGTGG

                852 accl, 859 fokl, 863 mnll, 871 accl hindll sall, 872 taql
                , 878 ecorl,
```

FIG. 12B

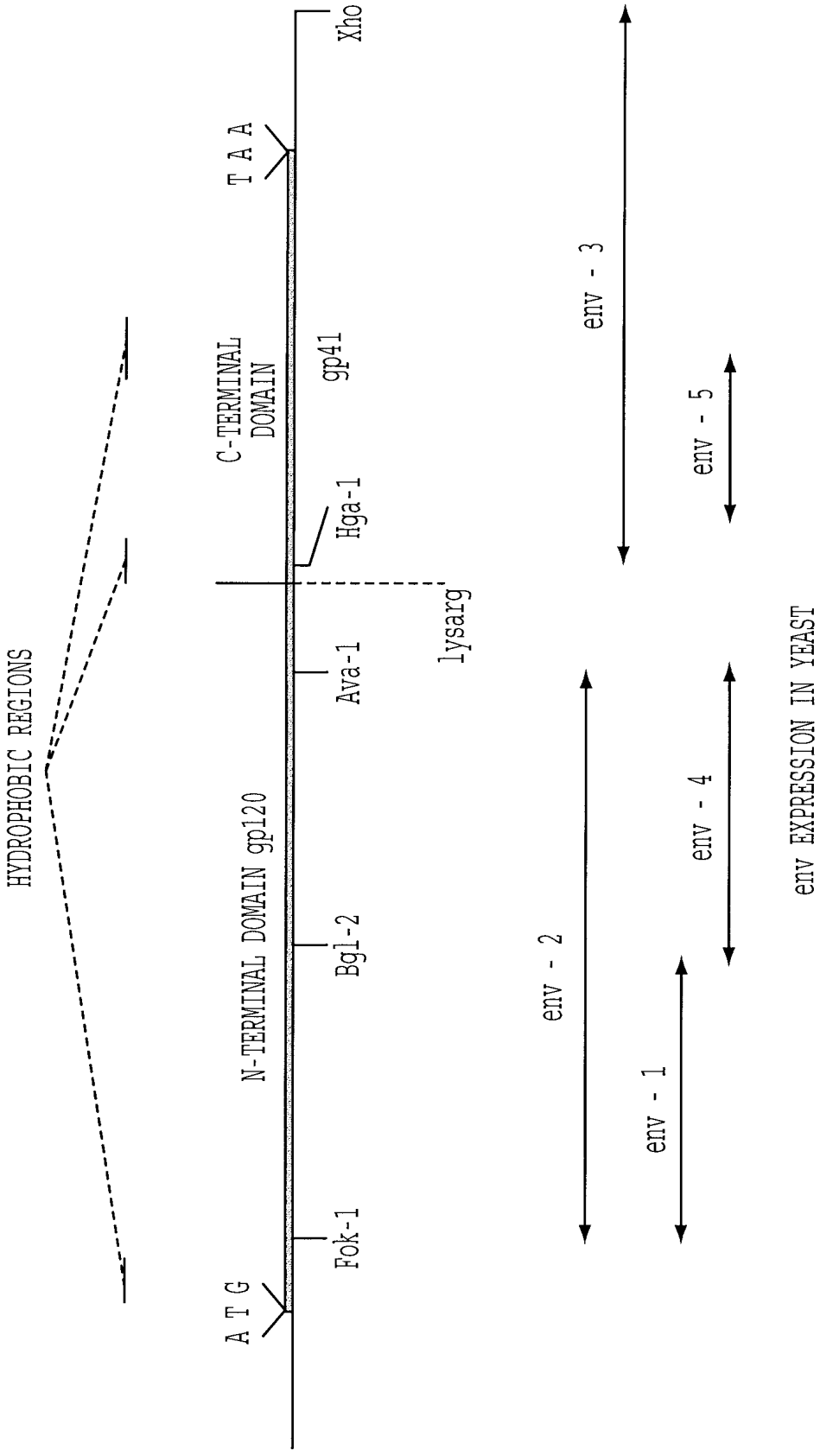

HYDROPHOBIC REGIONS
N-TERMINAL DOMAIN gp120
C-TERMINAL DOMAIN
gp41
A T G
T A A
Xho
Fok-1
Bgl-2
Ava-1
lysarg
Hga-1
env - 1
env - 2
env - 3
env - 4
env - 5
env EXPRESSION IN YEAST AND BACTERIA

FIG. 13

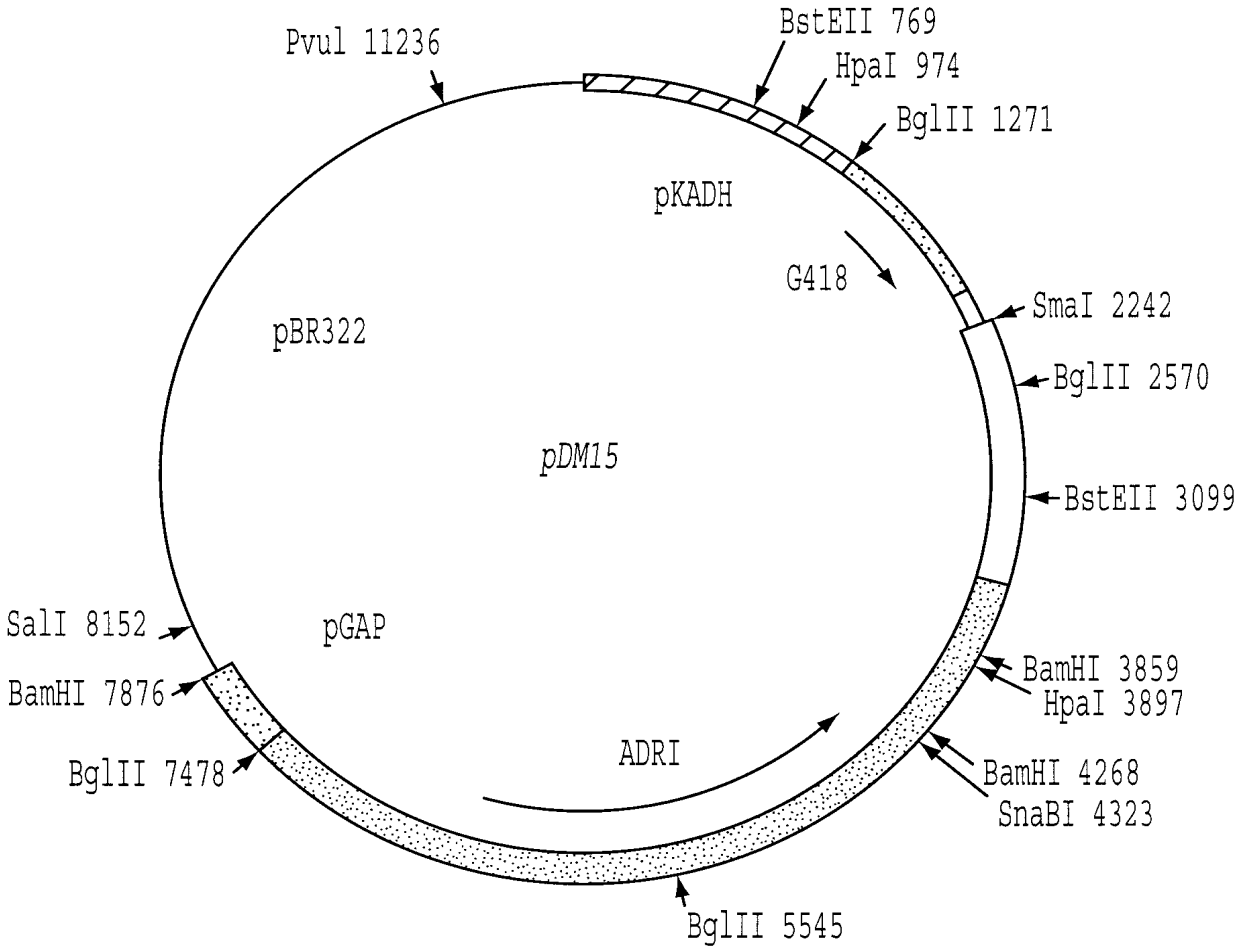
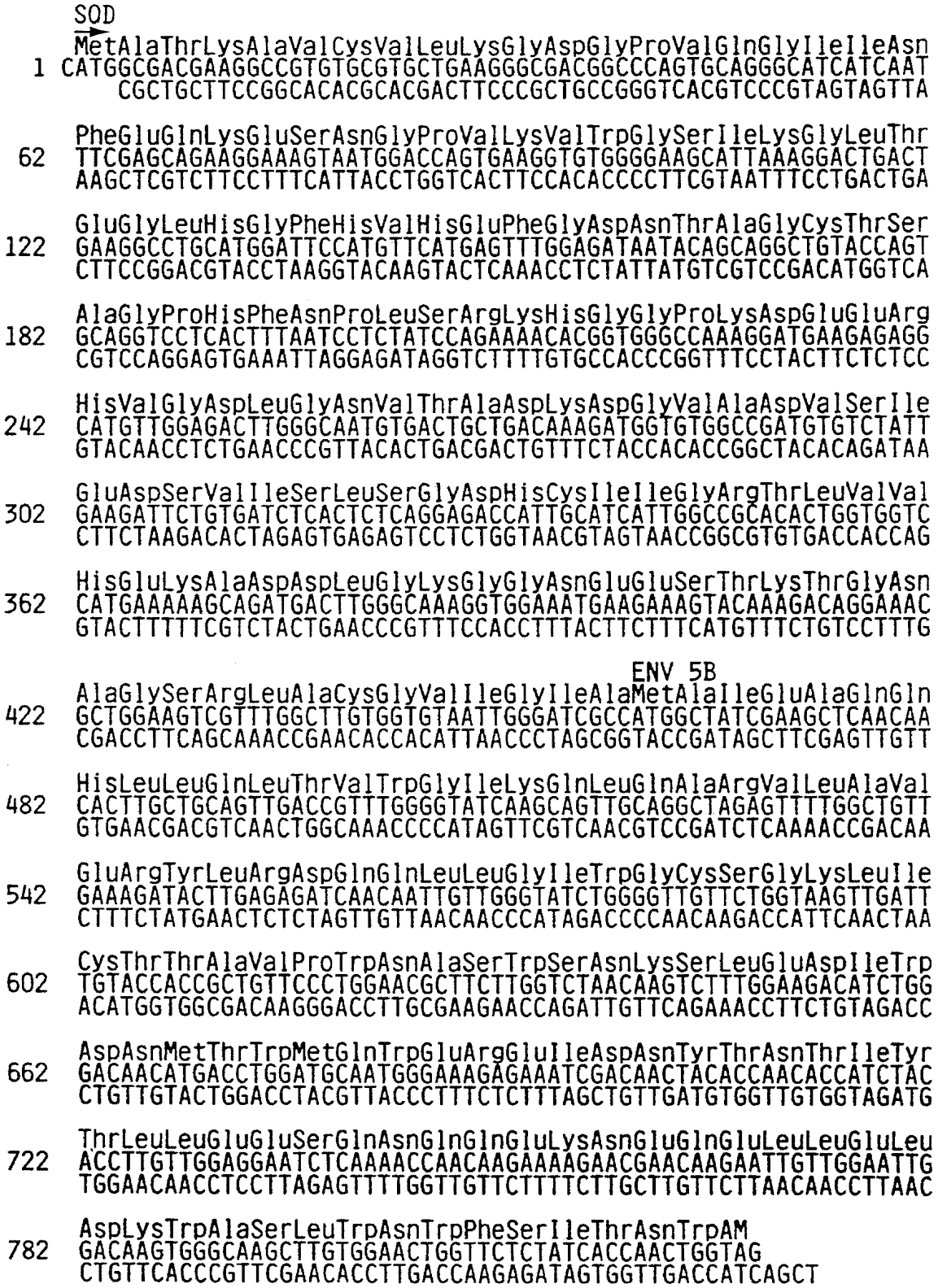

Pvul 11236
BstEII 769
HpaI 974
BglII 1271
pKADH
G418
SmaI 2242
BglII 2570
pBR322
pDM15
BstEII 3099
SalI 8152
pGAP
BamHI 7876
BamHI 3859
HpaI 3897
BglII 7478
ADRI
BamHI 4268
SnaBI 4323
BglII 5545

FIG. 14



Translated Mol. Weight = 30414.22

FIG. 15

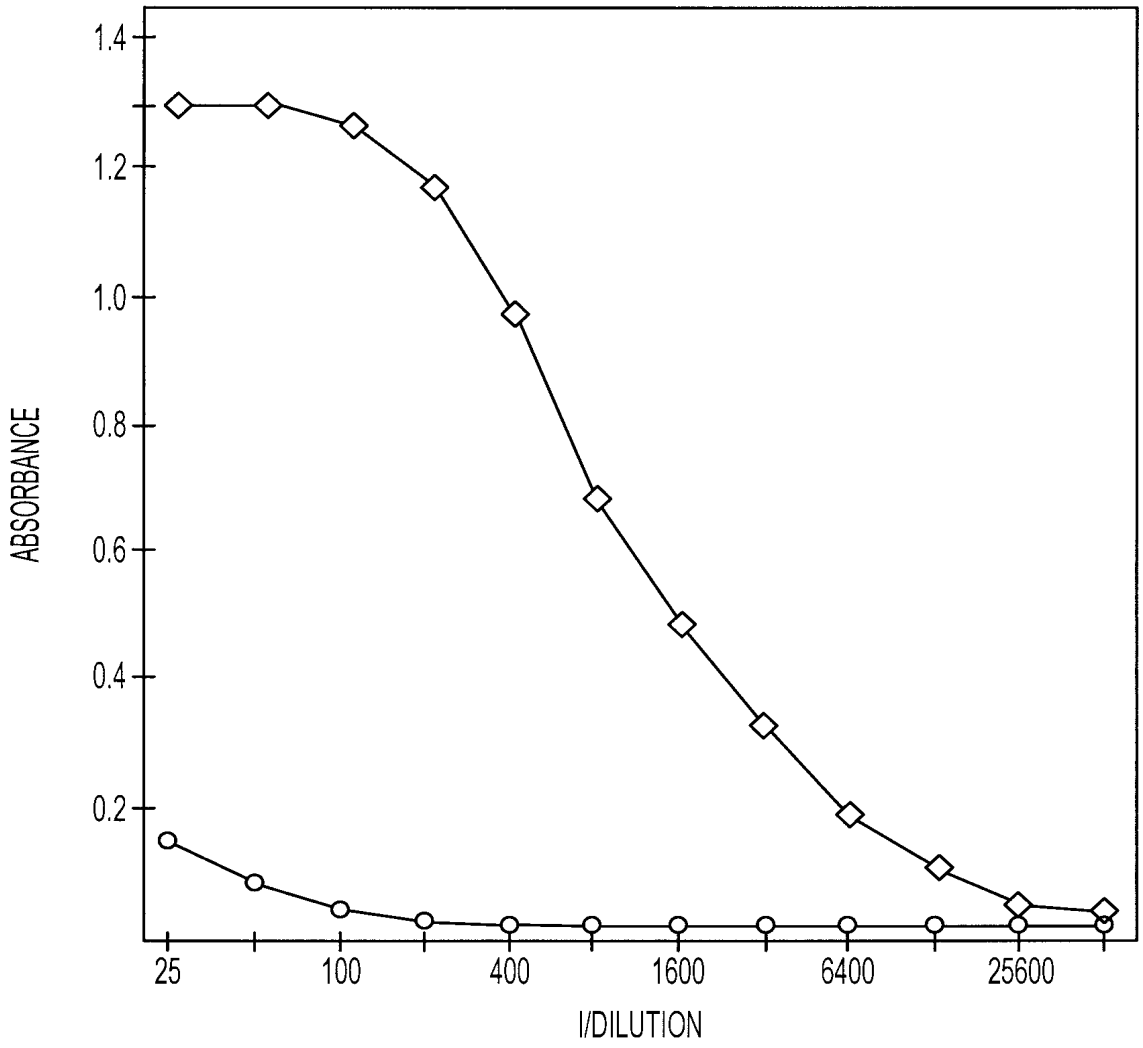

1.4
1.2
1.0
0.8
0.6
0.4
0.2
ABSORBANCE
25
100
400
1600
6400
25600
I/DILUTION

FIG. 16A

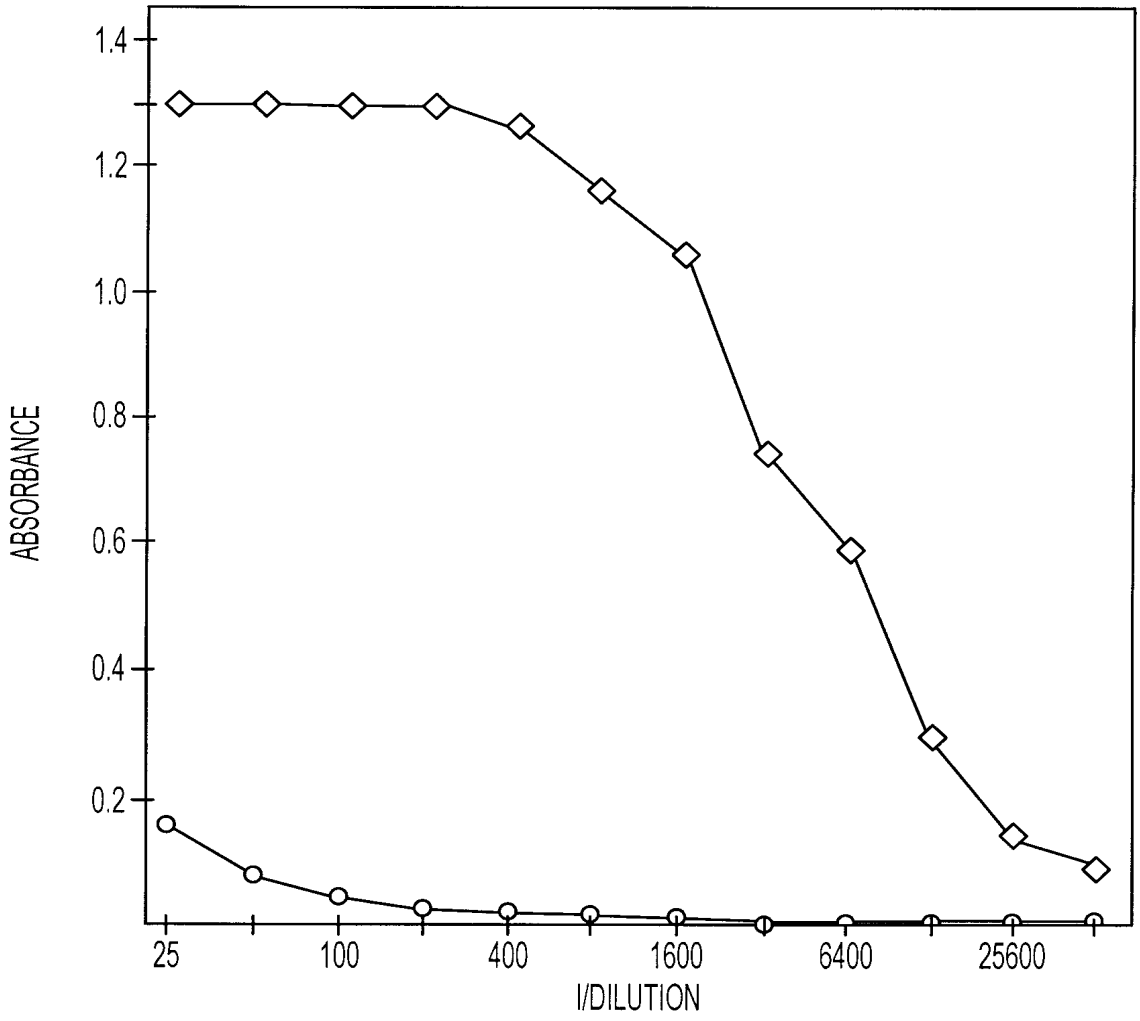

1.4
1.2
1.0
0.8
0.6
0.4
0.2
ABSORBANCE
25
100
400
1600
6400
25600
I/DILUTION

FIG. 16B

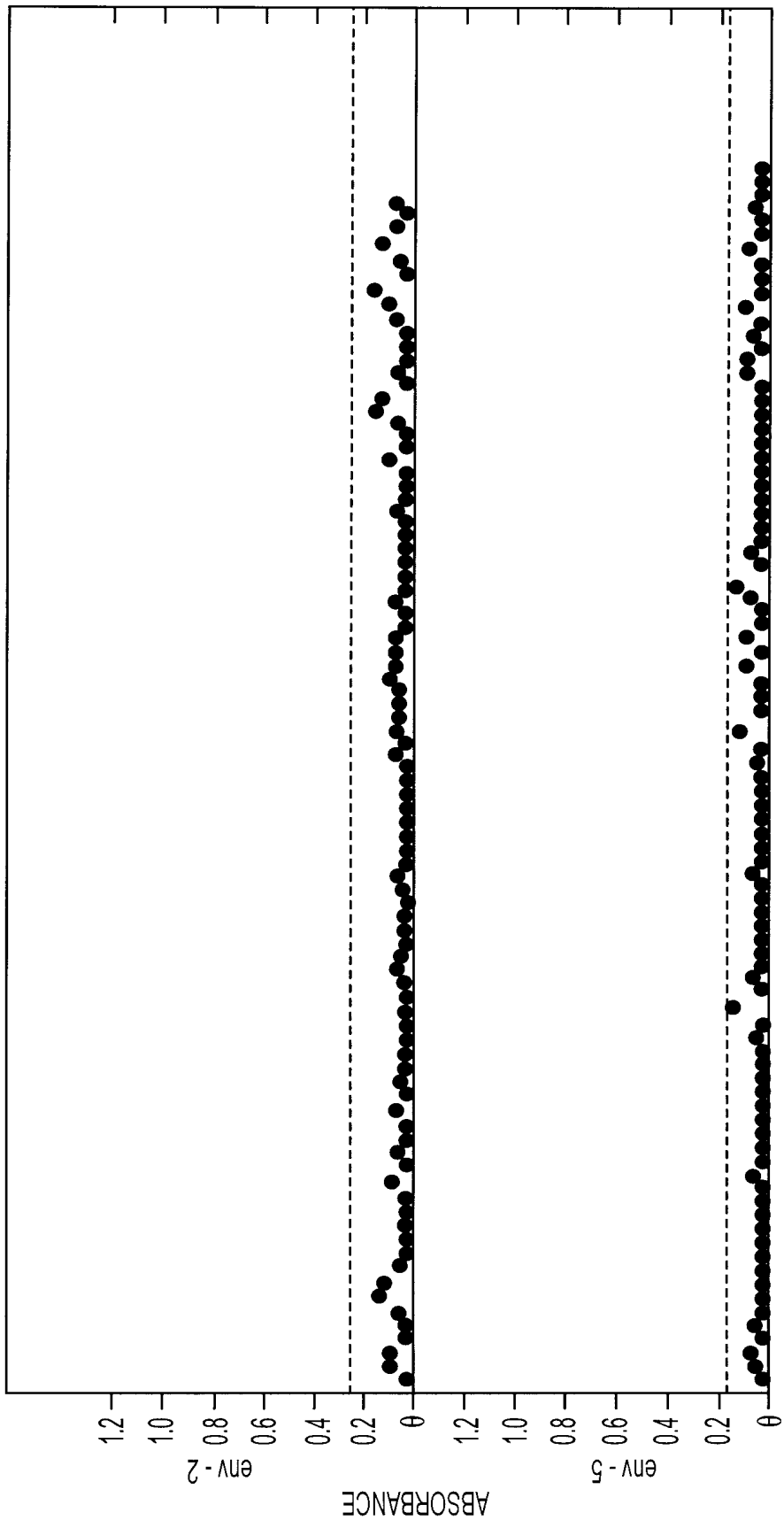


FIG. 17

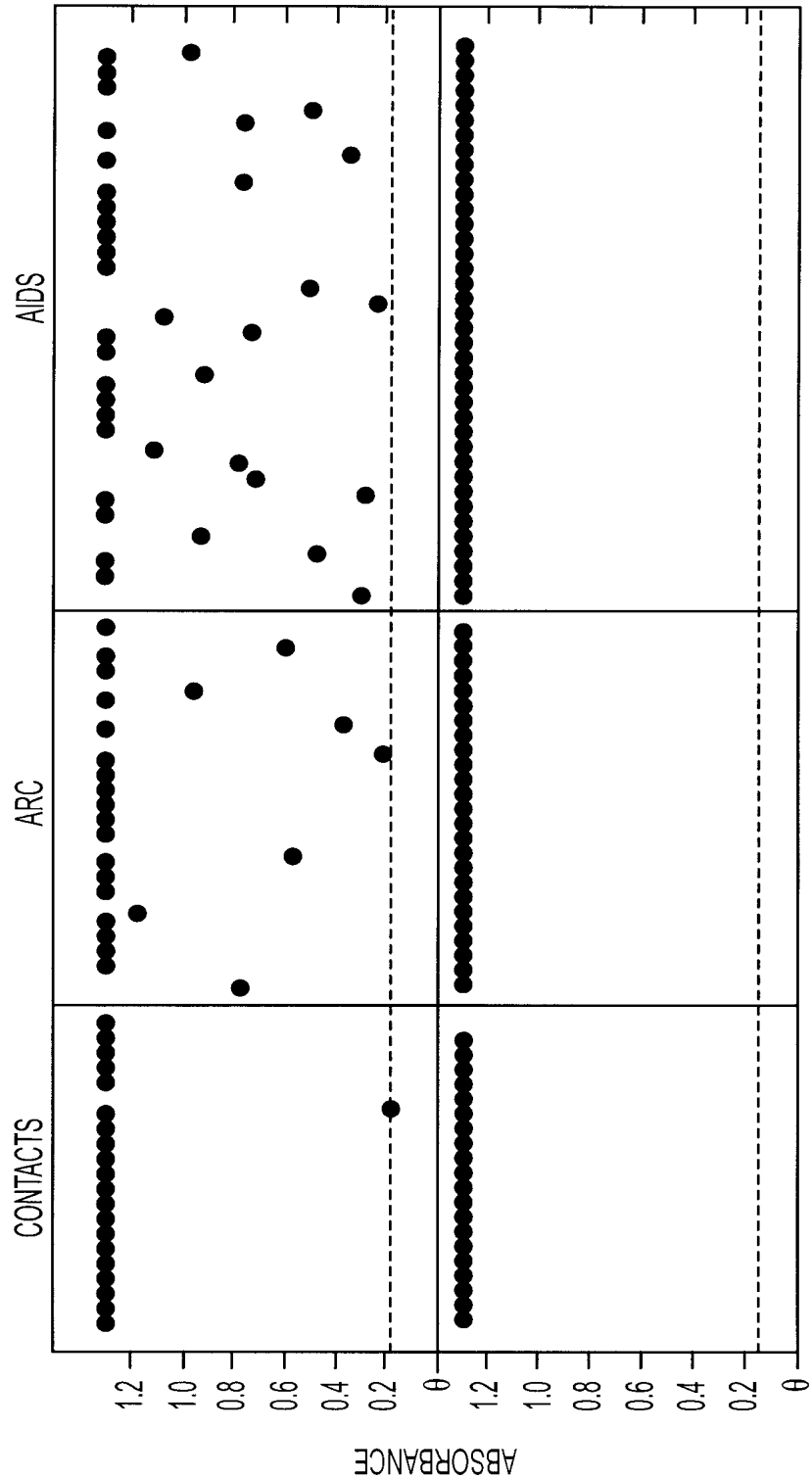


FIG. 18

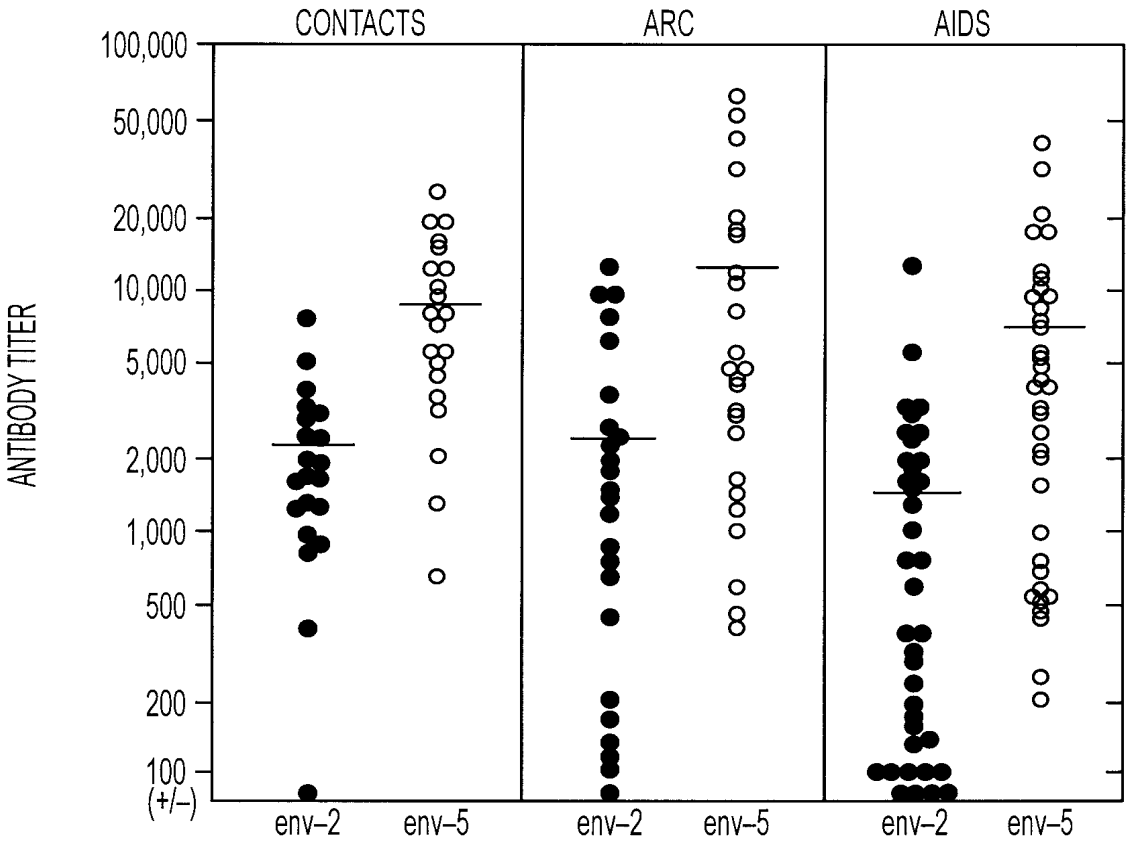

CONTACTS
ARC
AIDS
ANTIBODY TITER
100,000
50,000
20,000
10,000
5,000
2,000
1,000
500
200
100
(+/–)
env–2
env–5
env–2
env–5
env–2
env–5

FIG. 19

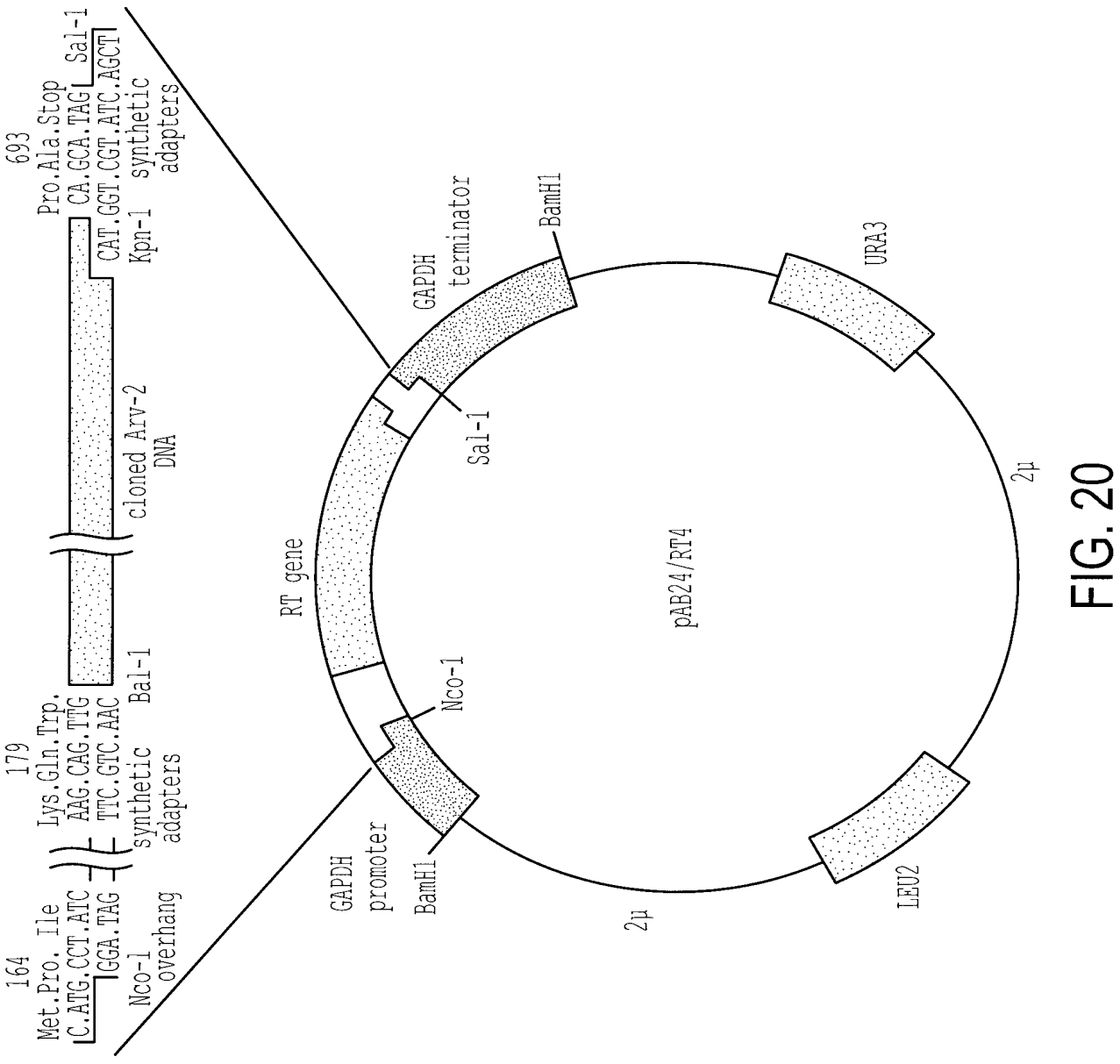

164
Met.Pro. Ile
C.ATG.CCT.ATC
TGA.TAG
Nco-1
overhang
179
Lys.Gln.Trp.
AAG.CAG.TTG
TTC.GTC.AAC
synthetic
adapters
Bal-1
cloned Arv-2
DNA
693
Pro.Ala.Stop
CA.GCA.TAG
CAT.GGT.CGT.ATC.AGCT
Kpn-1
synthetic
adapters
Sal-1
RT gene
GAPDH
promoter
BamH1
Nco-1
Sal-1
GAPDH
terminator
BamH1
pAB24/RT4
URA3
2μ
LEU2
2μ

FIG. 20

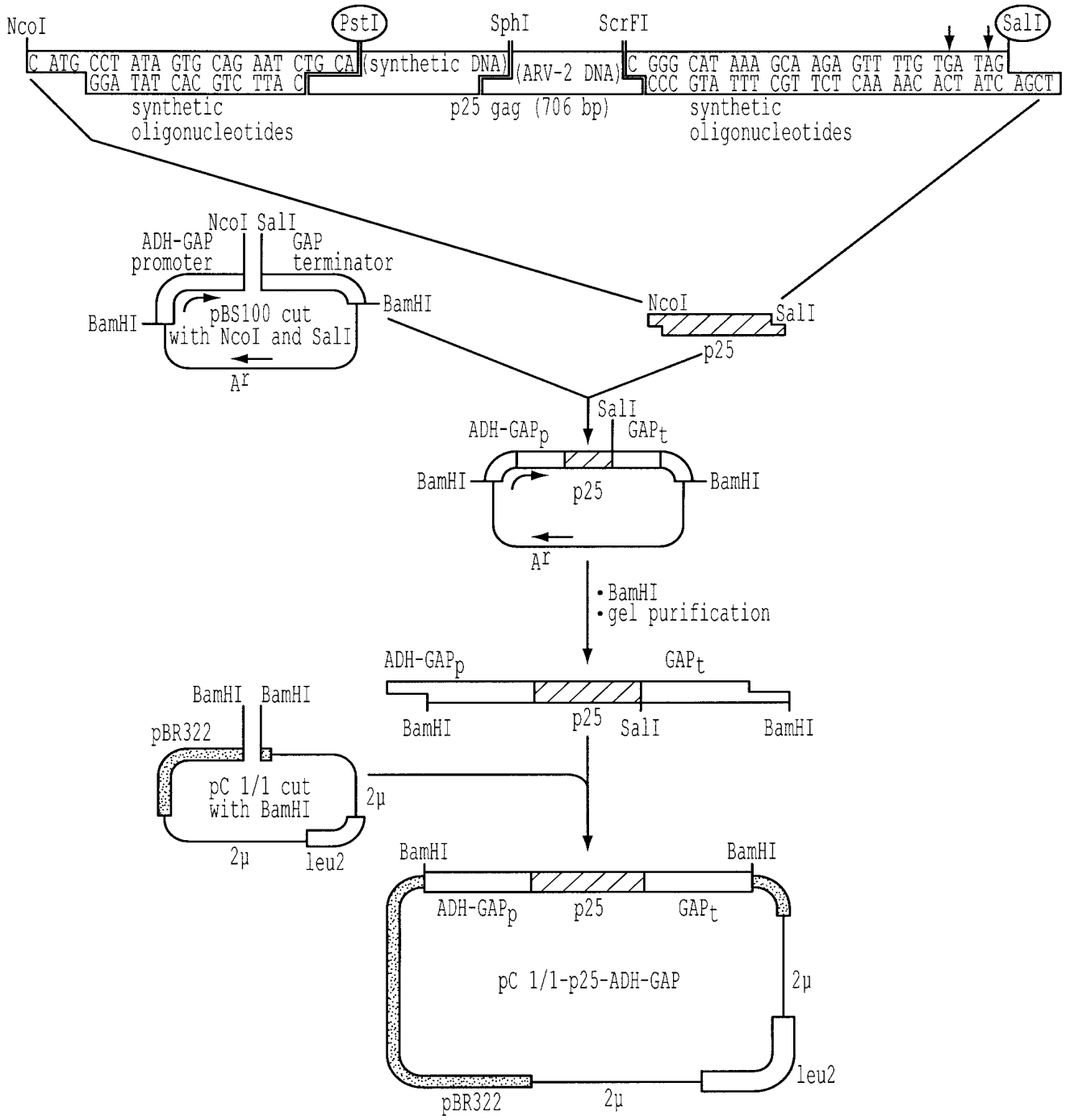

NcoI
PstI
SphI
ScrFI
SalI
C ATG CCT ATA GTG CAG AAT CTG CA
GGA TAT CAC GTC TTA C
(synthetic DNA)
(ARV-2 DNA)
C GGG CAT AAA GCA AGA GTT TTG TGA TAG
CCC GTA TTT CGT TCT CAA AAC ACT ATC AGCT
synthetic oligonucleotides
p25 gag (706 bp)
synthetic oligonucleotides
NcoI SalI
ADH-GAP promoter
GAP terminator
BamHI
pBS100 cut with NcoI and SalI
BamHI
Ar
NcoI
SalI
p25
SalI
ADH-GAPp
GAPt
BamHI
p25
BamHI
Ar
•BamHI
•gel purification
ADH-GAPp
GAPt
BamHI
p25
SalI
BamHI
BamHI BamHI
pBR322
pC 1/1 cut with BamHI
2μ
2μ
leu2
BamHI
BamHI
ADH-GAPp
p25
GAPt
pC 1/1-p25-ADH-GAP
2μ
leu2
pBR322
2μ

FIG. 21

```
      1                                                  10
     Met Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His Gln
C    ATG CCT ATA GTG CAG AAT CTG CAG GGG CAA ATG GTA CAT CAG

                        20
Ala Ile Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Val Glu
GCC ATA TCA CCT AGA ACT TTA AAT GCT TGG GTA AAA GTA GTA GAA

 30                                                 40
Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Ser Ala Leu
GAA AAG GCT TTC AGC CCA GAA GTA ATA CCC ATG TTT TCA GCA TTA

                        50
Ser Glu Gly Ala Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr
TCA GAA GGA GCC ACC CCT CAA GAT TTA AAC ACC ATG CTA AAC ACA

 60                                                 70
Val Gly Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile
GTG GGG GGA CAT CAA GCA GCC ATG CAA ATG TTA AAA GAG ACT ATC

                        80
Asn Glu Glu Ala Ala Glu Trp Asp Arg Val His Pro Val His Ala
AAT GAG GAG GCT GCC GAA TGG GAT AGA GTG CAT CCA GTG CAT GCA

 90                                                100
Gly Pro Ile Ala Pro Gly Gln Met Arg Glu Pro Arg Gly Ser Asp
GGG CCT ATT GCA CCA GGC CAA ATG AGA GAA CCA AGG GGA AGT GAC
```

FIG. 22A

```
                        110
Ile Ala Gly Thr Thr Ser Thr Leu Gln Glu Gln Ile Gly Trp Met
ATA GCA GGA ACT ACT AGT ACC CTT CAG GAA CAA ATA GGA TGG ATG

120                                             130
Thr Asn Asn Pro Pro Ile Pro Val Gly Glu Ile Tyr Lys Arg Trp
ACA AAT AAT CCA CCT ATC CCA GTA GGA GAA ATC TAT AAA AGA TGG

                        140
Ile Ile Leu Gly Leu Asn Lys Ile Val Arg Met Tyr Ser Pro Thr
ATA ATC CTG GGA TTA AAT AAA ATA GTA AGA ATG TAT AGC CCT ACC

150                                             160
Ser Ile Leu Asp Ile Arg Gln Gly Pro Lys Glu Pro Phe Arg Asp
AGC ATT CTG GAC ATA AGA CAA GGA CCA AAG GAA CCC TTT AGA GAT

                        170
Tyr Val Asp Arg Phe Tyr Lys Thr Leu Arg Ala Glu Gln Ala Ser
TAT GTA GAC CGG TTC TAT AAA ACT CTA AGA GCC GAA CAA GCT TCA

180                                             190
Gln Asp Val Lys Asn Trp Met Thr Glu Thr Leu Leu Val Gln Asn
CAG GAT GTA AAA AAT TGG ATG ACA GAA ACC TTG TTG GTC CAA AAT

                        200
Ala Asn Pro Asp Cys Lys Thr Ile Leu Lys Ala Leu Gly Pro Ala
GCA AAC CCA GAT TGT AAG ACT ATT TTA AAA GCA TTG GGA CCA GCA

210                                             220
Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly Val Gly Gly
GCT ACA CTA GAA GAA ATG ATG ACA GCA TGT CAG GGA GTG GGG GGA

                        230     232
Pro Gly His Lys Ala Arg Val Leu OP
CCC GGG CAT AAA GCA AGA GTT TTG TGA TAG
```

Translated Mol. Weight = 25700.75

FIG. 22B

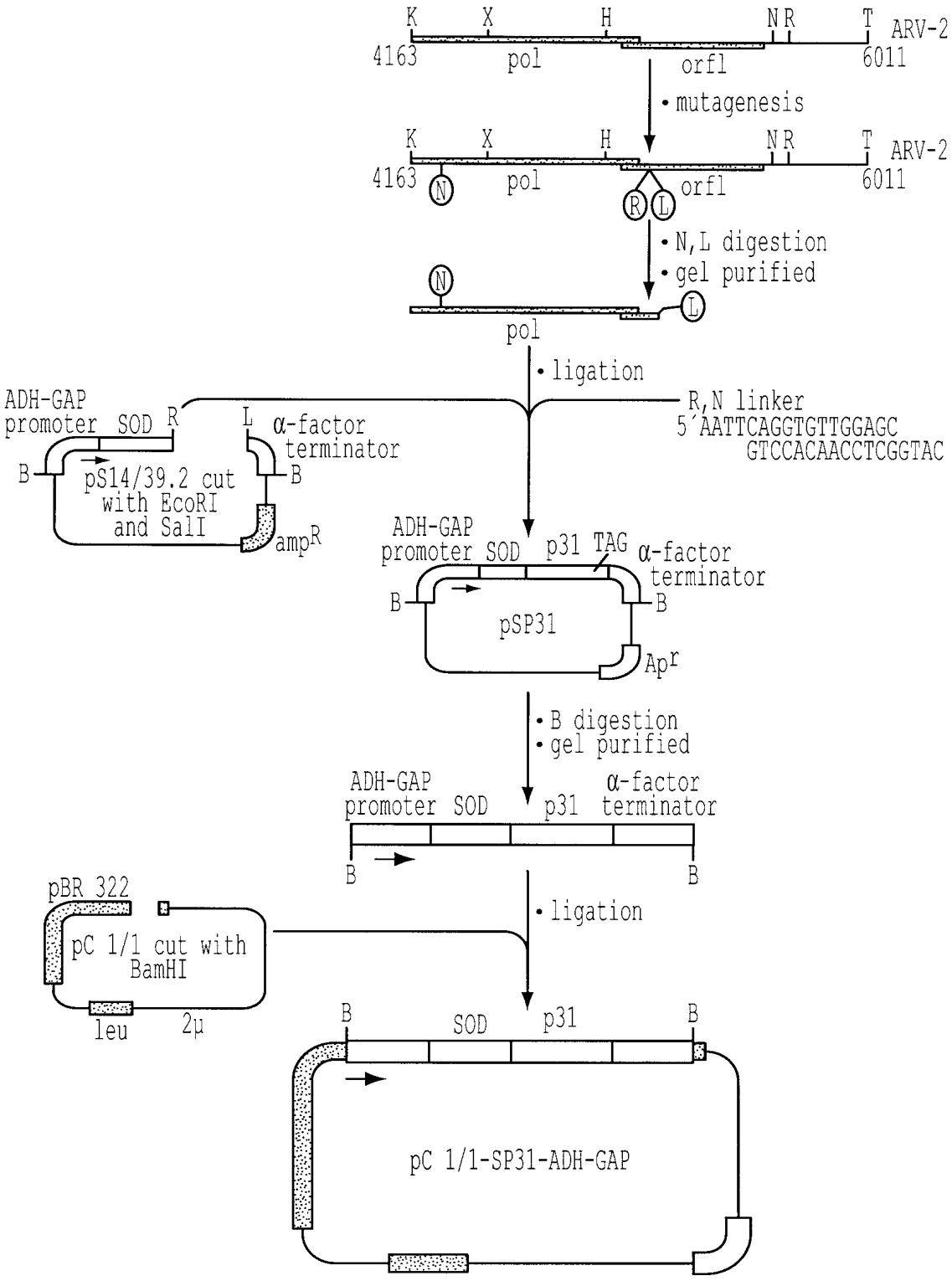

K
X
H
N R
T
ARV-2
4163
pol
orfl
6011
• mutagenesis
N
R
L
• N,L digestion
• gel purified
• ligation
ADH-GAP promoter
SOD
α-factor terminator
B
pS14/39.2 cut with EcoRI and SalI
ampR
R,N linker
5´AATTCAGGTGTTGGAGC
GTCCACAACCTCGGTAC
p31
TAG
pSP31
Apr
• B digestion
• gel purified
pBR 322
pC 1/1 cut with BamHI
leu
2μ
pC 1/1-SP31-ADH-GAP

FIG. 23

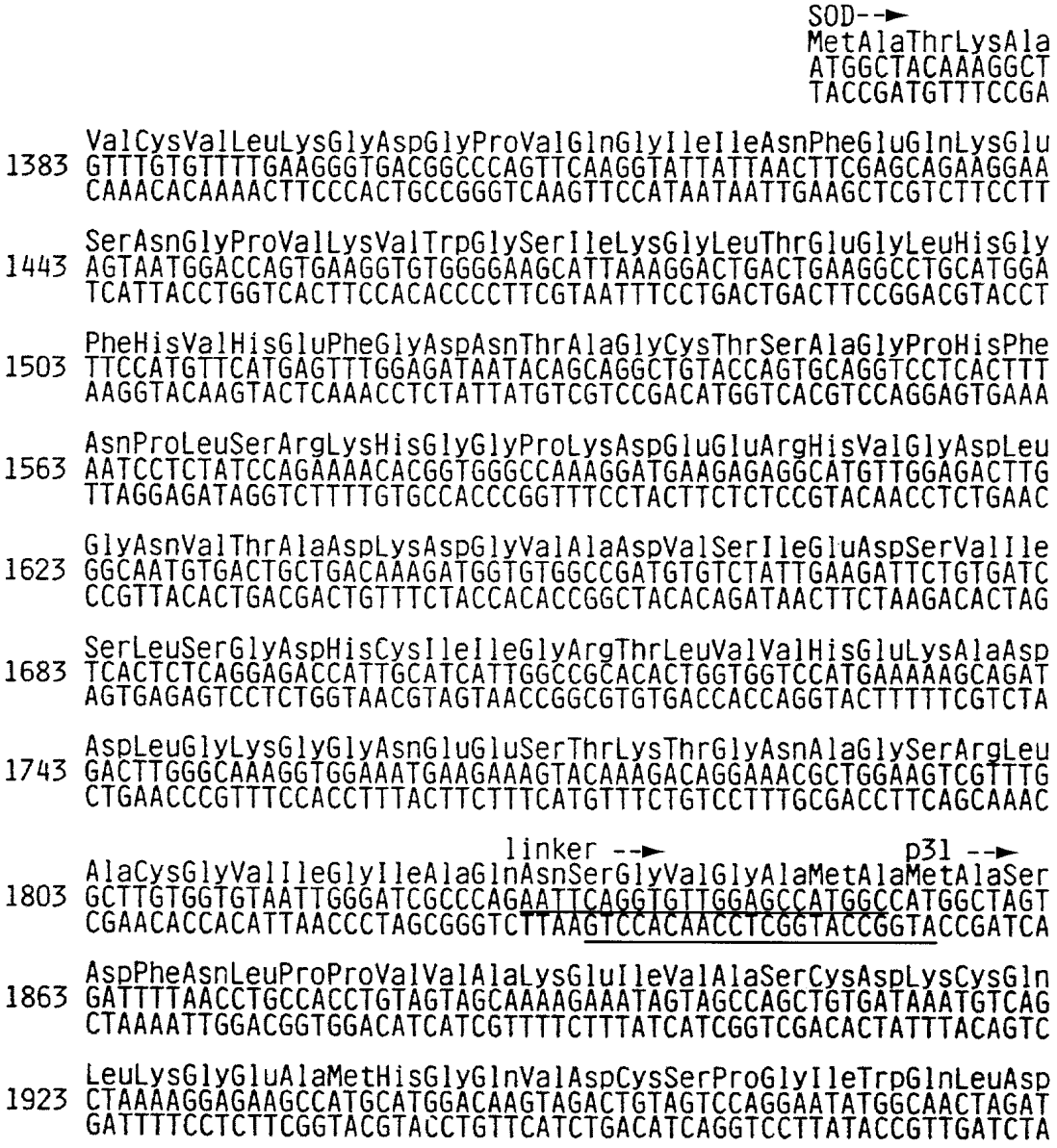

SOD-->
MetAlaThrLysAla
ATGGCTACAAAGGCT
TACCGATGTTTCCGA
ValCysValLeuLysGlyAspGlyProValGlnGlyIleIleAsnPheGluGlnLysGlu
1383 GTTTGTGTTTTGAAGGGTGACGGCCCAGTTCAAGGTATTATTAACTTCGAGCAGAAGGAA
CAAACACAAAACTTCCCACTGCCGGGTCAAGTTCCATAATAATTGAAGCTCGTCTTCCTT
SerAsnGlyProValLysValTrpGlySerIleLysGlyLeuThrGluGlyLeuHisGly
1443 AGTAATGGACCAGTGAAGGTGTGGGGAAGCATTAAAGGACTGACTGAAGGCCTGCATGGA
TCATTACCTGGTCACTTCCACACCCCTTCGTAATTTCCTGACTGACTTCCGGACGTACCT
PheHisValHisGluPheGlyAspAsnThrAlaGlyCysThrSerAlaGlyProHisPhe
1503 TTCCATGTTCATGAGTTTGGAGATAATACAGCAGGCTGTACCAGTGCAGGTCCTCACTTT
AAGGTACAAGTACTCAAACCTCTATTATGTCGTCCGACATGGTCACGTCCAGGAGTGAAA
AsnProLeuSerArgLysHisGlyGlyProLysAspGluGluArgHisValGlyAspLeu
1563 AATCCTCTATCCAGAAAACACGGTGGGCCAAAGGATGAAGAGAGGCATGTTGGAGACTTG
TTAGGAGATAGGTCTTTTGTGCCACCCGGTTTCCTACTTCTCTCCGTACAACCTCTGAAC
GlyAsnValThrAlaAspLysAspGlyValAlaAspValSerIleGluAspSerValIle
1623 GGCAATGTGACTGCTGACAAAGATGGTGTGGCCGATGTGTCTATTGAAGATTCTGTGATC
CCGTTACACTGACGACTGTTTCTACCACACCGGCTACACAGATAACTTCTAAGACACTAG
SerLeuSerGlyAspHisCysIleIleGlyArgThrLeuValValHisGluLysAlaAsp
1683 TCACTCTCAGGAGACCATTGCATCATTGGCCGCACACTGGTGGTCCATGAAAAAGCAGAT
AGTGAGAGTCCTCTGGTAACGTAGTAACCGGCGTGTGACCACCAGGTACTTTTTCGTCTA
AspLeuGlyLysGlyGlyAsnGluGluSerThrLysThrGlyAsnAlaGlySerArgLeu
1743 GACTTGGGCAAAGGTGGAAATGAAGAAAGTACAAAGACAGGAAACGCTGGAAGTCGTTTG
CTGAACCCGTTTCCACCTTTACTTCTTTCATGTTTCTGTCCTTTGCGACCTTCAGCAAAC
linker -->
p31 -->
AlaCysGlyValIleGlyIleAlaGlnAsnSerGlyValGlyAlaMetAlaMetAlaSer
1803 GCTTGTGGTGTAATTGGGATCGCCCAGAATTCAGGTGTTGGAGCCATGGCCATGGCTAGT
CGAACACCACATTAACCCTAGCGGGTCTTAAGTCCACAACCTCGGTACCGGTACCGATCA
AspPheAsnLeuProProValValAlaLysGluIleValAlaSerCysAspLysCysGln
1863 GATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAG
CTAAAATTGGACGGTGGACATCATCGTTTTCTTTATCATCGGTCGACACTATTTACAGTC
LeuLysGlyGluAlaMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAsp
1923 CTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGAT
GATTTTCCTCTTCGGTACGTACCTGTTCATCTGACATCAGGTCCTTATACCGTTGATCTA

FIG. 24A

```
                CysThrHisLeuGluGlyLysIleIleLeuValAlaValHisValAlaSerGlyTyrIle
1983            TGTACACATCTAGAAGGAAAAATTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATA
                ACATGTGTAGATCTTCCTTTTTTAATAGGACCATCGTCAAGTACATCGGTCACCTATATAT

                GluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLeuLysLeu
2043            GAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAAACAGCATATTTTCTCTTAAAATTA
                CTTCGTCTTCAATAAGGTCGTCTCTGTCCCGTCCTTTGTCGTATAAAAGAGAATTTTAAT

                AlaGlyArgTrpProValLysThrIleHisThrAspAsnGlySerAsnPheThrSerThr
2103            GCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCAGCAATTTCACCAGTACT
                CGTCCTTCTACCGGTCATTTTTGTTATGTATGTCTGTTACCGTCGTTAAAGTGGTCATGA

                ThrValLysAlaAlaCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsn
2163            ACGGTTAAGGCCGCCTGTTGGTGGGCAGGGATCAAGCAGGAATTTGGCATTCCCTACAAT
                TGCCAATTCCGGCGGACAACCACCCGTCCCTAGTTCGTCCTTAAACCGTAAGGGATGTTA

                ProGlnSerGlnGlyValValGluSerMetAsnAsnGluLeuLysLysIleIleGlyGln
2223            CCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAATGAATTAAAGAAAATTATAGGACAG
                GGGGTTTCAGTTCCTCATCATCTTAGATACTTATTACTTAATTTCTTTTAATATCCTGTC

                ValArgAspGlnAlaGluHisLeuLysThrAlaValGlnMetAlaValPheIleHisAsn
2283            GTAAGAGATCAGGCTGAACACCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAAT
                CATTCTCTAGTCCGACTTGTGGAATTCTGTCGTCATGTTTACCGTCATAAGTAGGTGTTA

                PheLysArgLysGlyGlyIleGlyGlyTyrSerAlaGlyGluArgIleValAspIleIle
2343            TTTAAAAGAAAAGGGGGGATTGGGGGATACAGTGCAGGGGAAAGAATAGTAGACATAATA
                AAATTTTCTTTTCCCCCCTAACCCCCTATGTCACGTCCCCTTTCTTATCATCTGTATTAT

                AlaThrAspIleGlnThrLysGluLeuGlnLysGlnIleThrLysIleGlnAsnPheArg
2403            GCAACAGACATACAAACTAAAGAACTACAAAAGCAAATTACAAAAATTCAAAATTTTCGG
                CGTTGTCTGTATGTTTGATTTCTTGATGTTTTCGTTTAATGTTTTTAAGTTTTAAAAGCC

                ValTyrTyrArgAspAsnLysAspProLeuTrpLysGlyProAlaLysLeuLeuTrpLys
2463            GTTTATTACAGGGACAACAAAGATCCCCTTTGGAAAGGACCAGCAAAGCTTCTCTGGAAA
                CAAATAATGTCCCTGTTGTTTCTAGGGGAAACCTTTCCTGGTCGTTTCGAAGAGACCTTT

                GlyGluGlyAlaValValIleGlnAspAsnSerAspIleLysValValProArgArgLys
2523            GGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAA
                CCACTTCCCCGTCATCATTATGTTCTATTATCACTGTATTTTCATCACGGTTCTTCTTTT

                AlaLysIleIleArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArg
2583            GCAAAAATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGA
                CGTTTTTAGTAATCCCTAATACCTTTTGTCTACCGTCCACTACTAACACACCGTTCATCT

                GlnAspGluAspAM
2643            CAGGATGAGGATTAG
                GTCCTACTCCTAATC
```

FIG. 24B

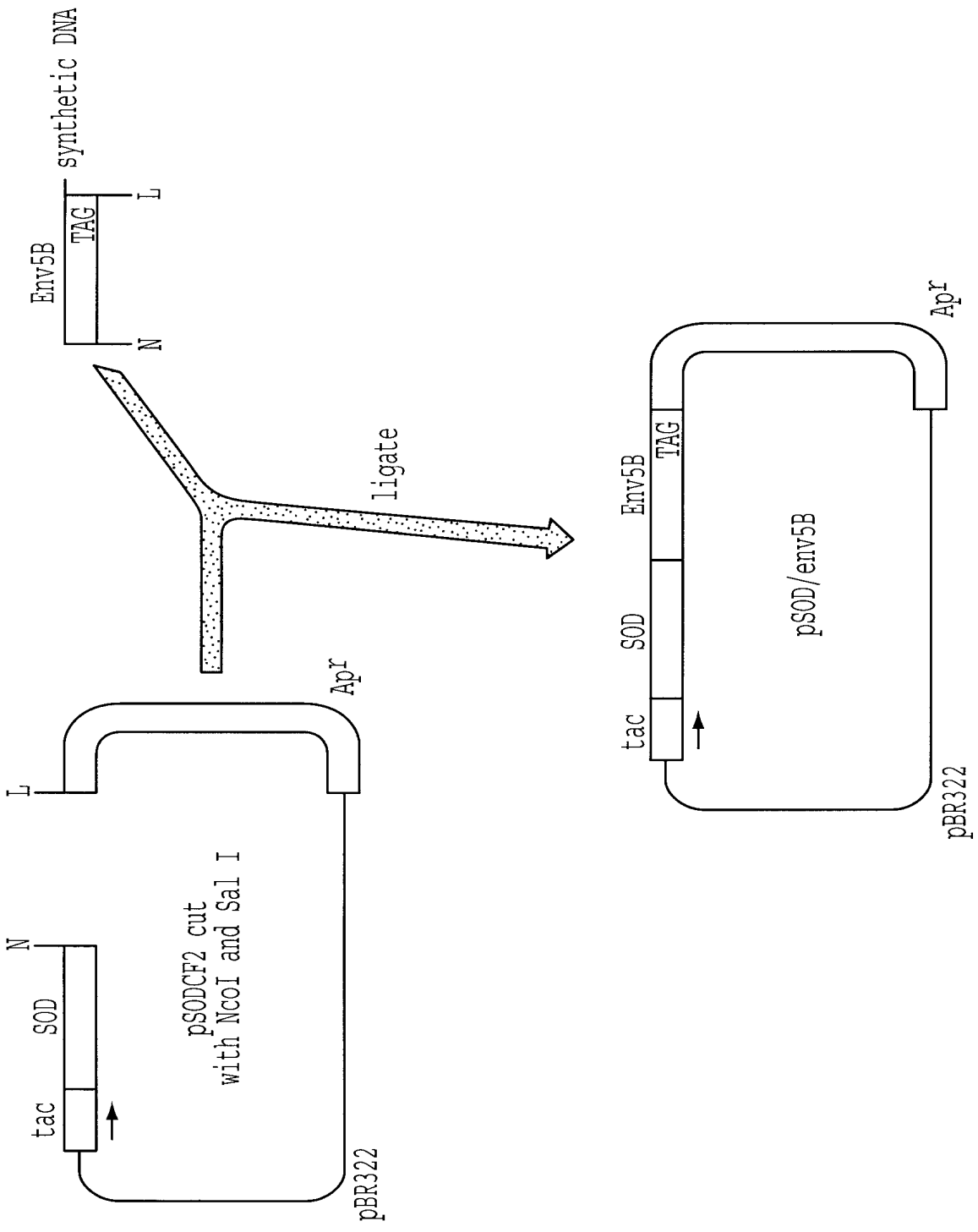

synthetic DNA
Env5B
TAG
L
N
ligate
pSODCF2 cut
with NcoI and Sal I
Ap^r
pBR322
tac
SOD
pSOD/env5B

FIG. 25

Sequence of SOD/env-4

```
     SOD --►
     MetAlaThrLysAlaValCysValLeuLysGlyAspGlyProValGlnGlyIleIleAsn
   1 CATGGCGACGAAGGCCGTGTGCGTGCTGAAGGGCGACGGCCCAGTGCAGGGCATCATCAAT
       CGCTGCTTCCGGCACACGCACGACTTCCCGCTGCCGGGTCACGTCCCGTAGTAGTTA

     PheGluGlnLysGluSerAsnGlyProValLysValTrpGlySerIleLysGlyLeuThr
  62 TTCGAGCAGAAGGAAAGTAATGGACCAGTGAAGGTGTGGGGAAGCATTAAAGGACTGACT
     AAGCTCGTCTTCCTTTCATTACCTGGTCACTTCCACACCCCTTCGTAATTTCCTGACTGA

     GluGlyLeuHisGlyPheHisValHisGluPheGlyAspAsnThrAlaGlyCysThrSer
 122 GAAGGCCTGCATGGATTCCATGTTCATGAGTTTGGAGATAATACAGCAGGCTGTACCAGT
     CTTCCGGACGTACCTAAGGTACAAGTACTCAAACCTCTATTATGTCGTCCGACATGGTCA

     AlaGlyProHisPheAsnProLeuSerArgLysHisGlyGlyProLysAspGluGluArg
 182 GCAGGTCCTCACTTTAATCCTCTATCCAGAAAACACGGTGGGCCAAAGGATGAAGAGAGG
     CGTCCAGGAGTGAAATTAGGAGATAGGTCTTTTGTGCCACCCGGTTTCCTACTTCTCTCC

     HisValGlyAspLeuGlyAsnValThrAlaAspLysAspGlyValAlaAspValSerIle
 242 CATGTTGGAGACTTGGGCAATGTGACTGCTGACAAAGATGGTGTGGCCGATGTGTCTATT
     GTACAACCTCTGAACCCGTTACACTGACGACTGTTTCTACCACACCGGCTACACAGATAA

     GluAspSerValIleSerLeuSerGlyAspHisCysIleIleGlyArgThrLeuValVal
 302 GAAGATTCTGTGATCTCACTCTCAGGAGACCATTGCATCATTGGCCGCACACTGGTGGTC
     CTTCTAAGACACTAGAGTGAGAGTCCTCTGGTAACGTAGTAACCGGCGTGTGACCACCAG

     HisGluLysAlaAspAspLeuGlyLysGlyGlyAsnGluGluSerThrLysThrGlyAsn
 362 CATGAAAAAGCAGATGACTTGGGCAAAGGTGGAAATGAAGAAAGTACAAAGACAGGAAAC
     GTACTTTTTCGTCTACTGAACCCGTTTCCACCTTTACTTCTTTCATGTTTCTGTCCTTTG

                                                    Env4--►
     AlaGlySerArgLeuAlaCysGlyValIleGlhIleAlaMetGluValValIleArgSer
 422 GCTGGAAGTCGTTTGGCTTGTGGTGTAATTGGGATCGCCATGGAGGTAGTAATTAGATCT
     CGACCTTCAGCAAACCGAACACCACATTAACCCTAGCGGTACCTCCATCATTAATCTAGA

     AspAsnPheThrAsnAsnAlaLysThrIleIleValGlnLeuAsnGluSerValAlaIle
 482 GACAATTTCACGAACAATGCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATT
     CTGTTAAAGTGCTTGTTACGATTTTGGTATTATCATGTCGACTTACTTAGACATCGTTAA

     AsnCysThrArgProAsnAsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAla
 542 AACTGTACAAGACCCAACAACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCA
     TTGACATGTTCTGGGTTGTTGTTATGTTCTTTTTCATAGATATATCCTGGTCCCTCTCGT
```

FIG. 26A

```
    PheHisThrThrGlyArgIleIleGlyAspIleArgLysAlaHisCysAsnIleSerArg
602 TTTCATACAACAGGAAGAATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGA
    AAAGTATGTTGTCCTTCTTATTATCCTCTATATTCTTTTCGTGTAACATTGTAATCATCT

    AlaGlnTrpAsnAsnThrLeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsn
662 GCACAATGGAATAACACTTTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAAT
    CGTGTTACCTTATTGTGAAATCTTGTCTATCAATTTTTTAATTCTCTTGTCAAACCCTTA

    AsnLysThrIleValPheAsnGlnSerSerGlyGlyAspProGluIleValMetHisSer
722 AATAAAACAATAGTCTTTAATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGT
    TTATTTTGTTATCAGAAATTAGTTAGGAGTCCTCCCCTGGGTCTTTAACATTACGTGTCA

    PheAsnCysArgGlyGluPhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrp
782 TTTAATTGTAGAGGGGAATTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGG
    AAATTAACATCTCCCCTTAAAAAGATGACATTATGTTGTGTTGACAAATTATTATGTACC

    ArgLeuAsnHisThrGluGlyThrLysGlyAsnAspThrIleIleLeuProCysArgIle
842 AGGTTAAATCACACTGAAGGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATA
    TCCAATTTAGTGTGACTTCCTTGATTTCCTTTACTGTGTTAGTATGAGGGTACATCTTAT

    LysGlnIleIleAsnMetTrpGlnGluValGlyLysAlaMetTyrAlaProProIleGly
902 AAACAAATTATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGA
    TTTGTTTAATATTTGTACACCGTCCTTCATCCTTTTCGTTACATACGGGGAGGGTAACCT

    GlyGlnIleSerCysSerSerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThr
962 GGACAAATTAGTTGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACA
    CCTGTTTAATCAACAAGTAGTTTATAATGTCCCGACGATAATTGTTCTCTACCACCATGT

    AsnValThrAsnAspThrGluValPheArgProGlyGlyGlyAspMetArgAspAsnTrp
1022 AATGTAACTAATGACACCGAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGG
     TTACATTGATTACTGTGGCTCCAGAAGTCTGGACCTCCTCCTCTATACTCCCTGTTAACC

     ArgSerGluLeuTyrLysTyrLysValIleLysIleGluProLeuGlyIleAlaProThr
1082 AGAAGTGAATTATATAAATATAAAGTAATAAAAATTGAACCATTAGGAATAGCACCCACC
     TCTTCACTTAATATATTTATATTTCATTATTTTTAACTTGGTAATCCTTATCGTGGGTGG

     LysAlaLysArgArgValValGlnArgGluLysArgOP OP
1142 AAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGATGATGAAGCTTG
     TTCCGTTTCTCTTCTCACCACGTCTCTCTTTTTTCTACTACTTCGAACAGCT
```

FIG. 26B

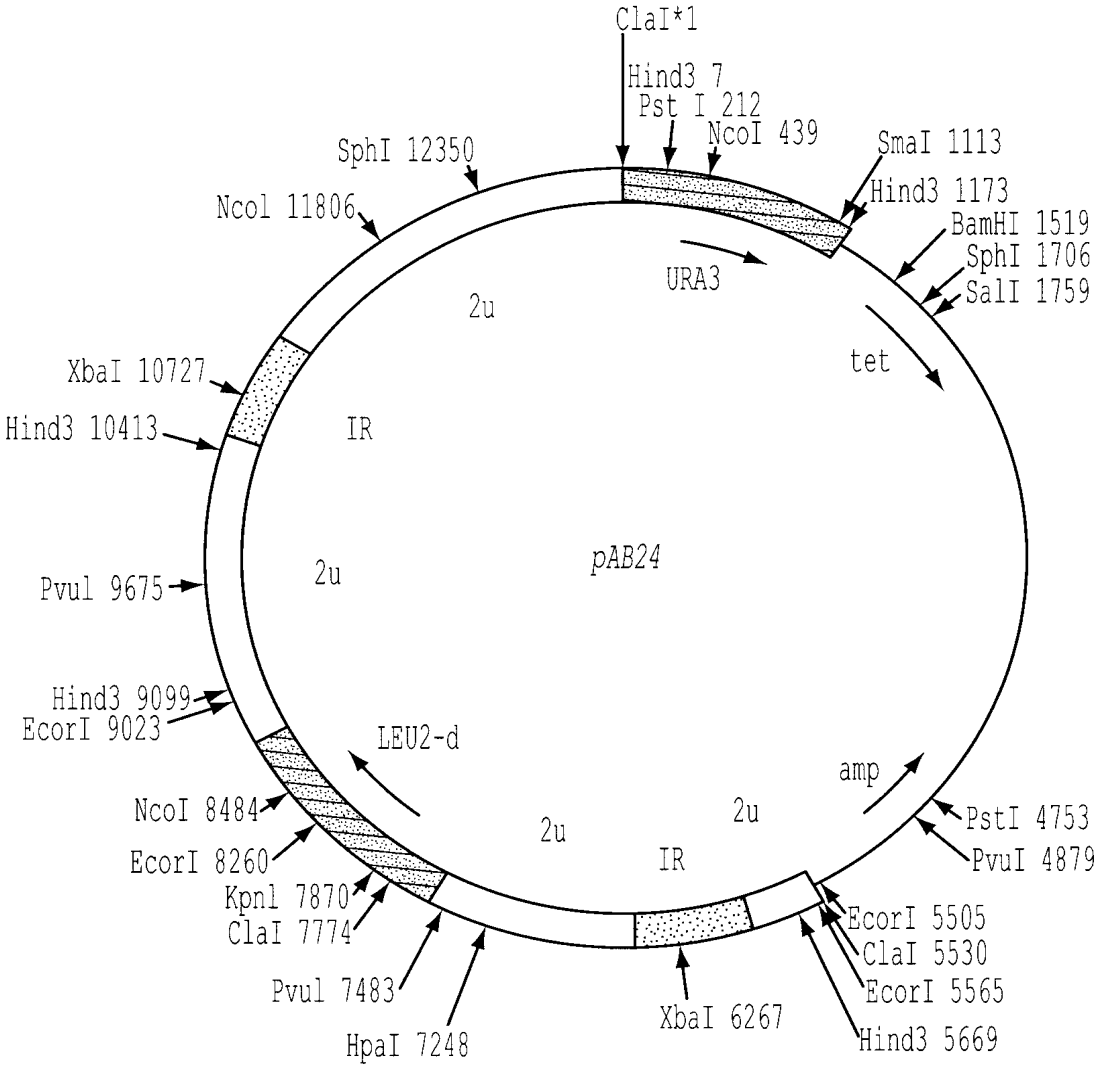

ClaI*1
Hind3 7
Pst I 212
NcoI 439
SmaI 1113
Hind3 1173
BamHI 1519
SphI 1706
SalI 1759
SphI 12350
Ncol 11806
URA3
2u
tet
XbaI 10727
Hind3 10413
IR
pAB24
2u
Pvul 9675
Hind3 9099
EcorI 9023
LEU2-d
amp
NcoI 8484
2u
2u
PstI 4753
EcorI 8260
IR
PvuI 4879
Kpnl 7870
ClaI 7774
EcorI 5505
ClaI 5530
EcorI 5565
Pvul 7483
XbaI 6267
Hind3 5669
HpaI 7248

FIG. 27

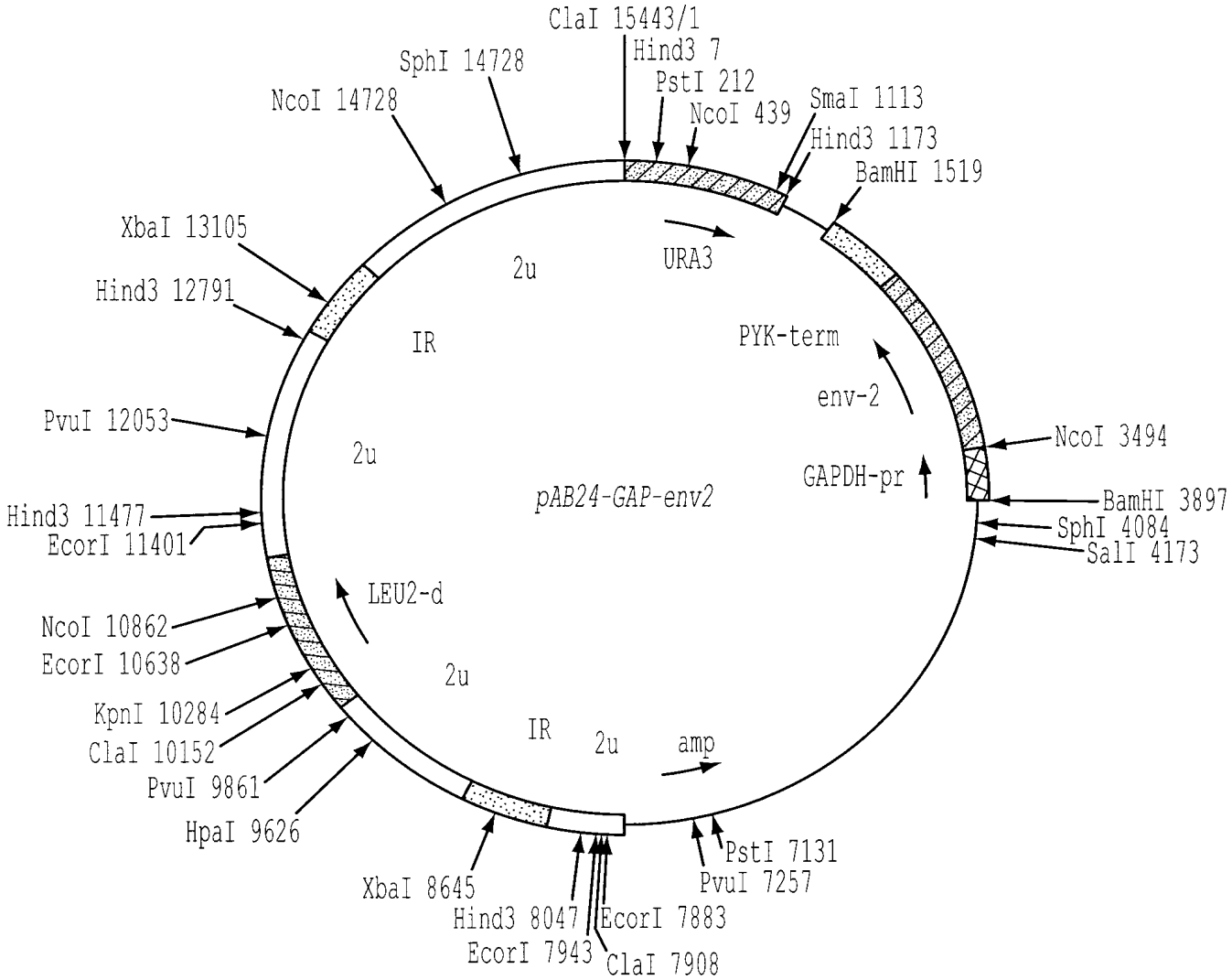

ClaI 15443/1
Hind3 7
PstI 212
NcoI 439
SmaI 1113
Hind3 1173
BamHI 1519
SphI 14728
NcoI 14728
XbaI 13105
Hind3 12791
PvuI 12053
Hind3 11477
EcorI 11401
NcoI 10862
EcorI 10638
KpnI 10284
ClaI 10152
PvuI 9861
HpaI 9626
XbaI 8645
Hind3 8047
EcorI 7943
EcorI 7883
ClaI 7908
PvuI 7257
PstI 7131
NcoI 3494
BamHI 3897
SphI 4084
SalI 4173
2u
URA3
IR
PYK-term
env-2
2u
GAPDH-pr
pAB24-GAP-env2
LEU2-d
2u
IR
2u
amp

FIG. 28

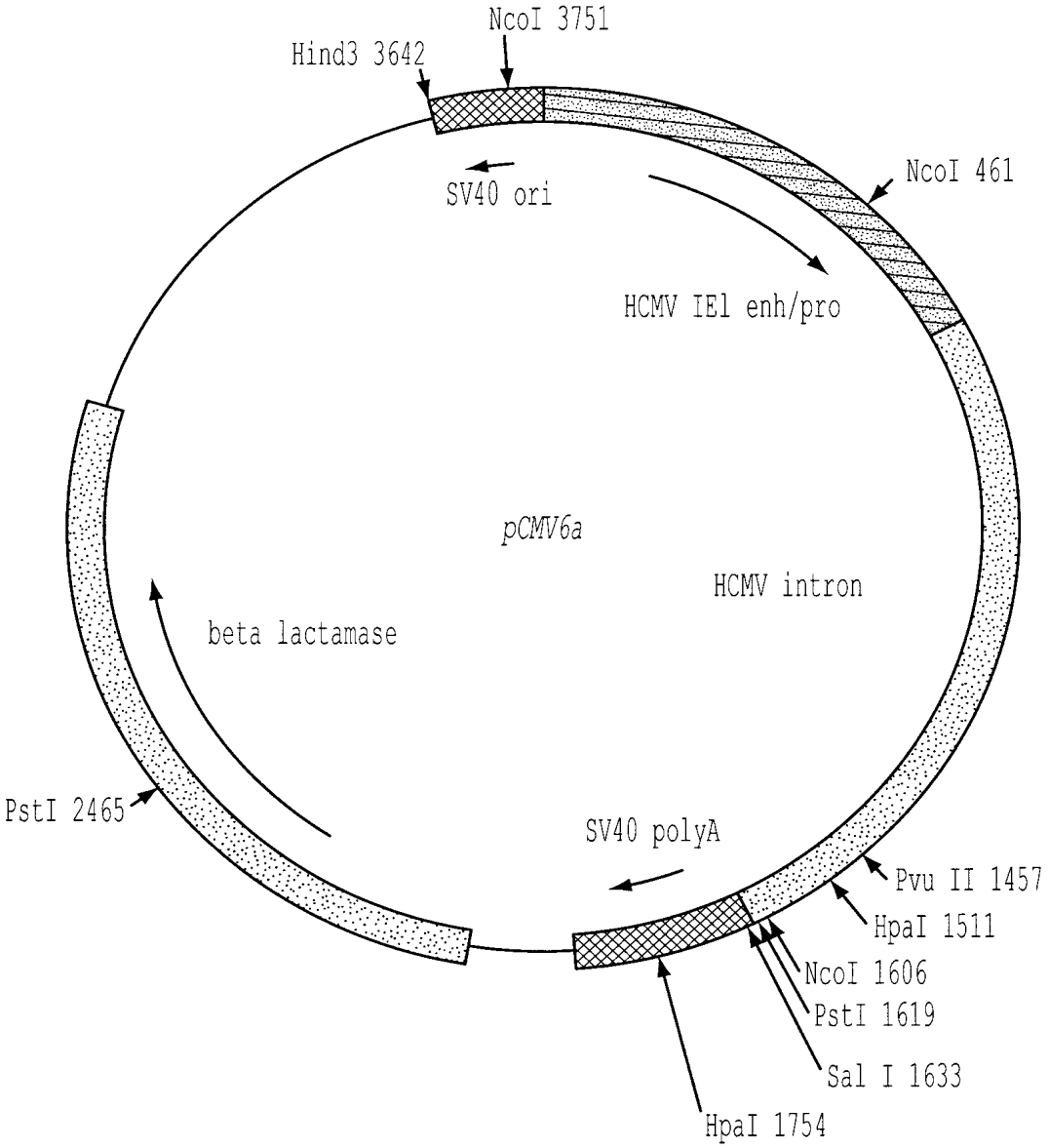

NcoI 3751
Hind3 3642
NcoI 461
SV40 ori
HCMV IEl enh/pro
pCMV6a
HCMV intron
beta lactamase
PstI 2465
SV40 polyA
Pvu II 1457
HpaI 1511
NcoI 1606
PstI 1619
Sal I 1633
HpaI 1754

FIG. 29

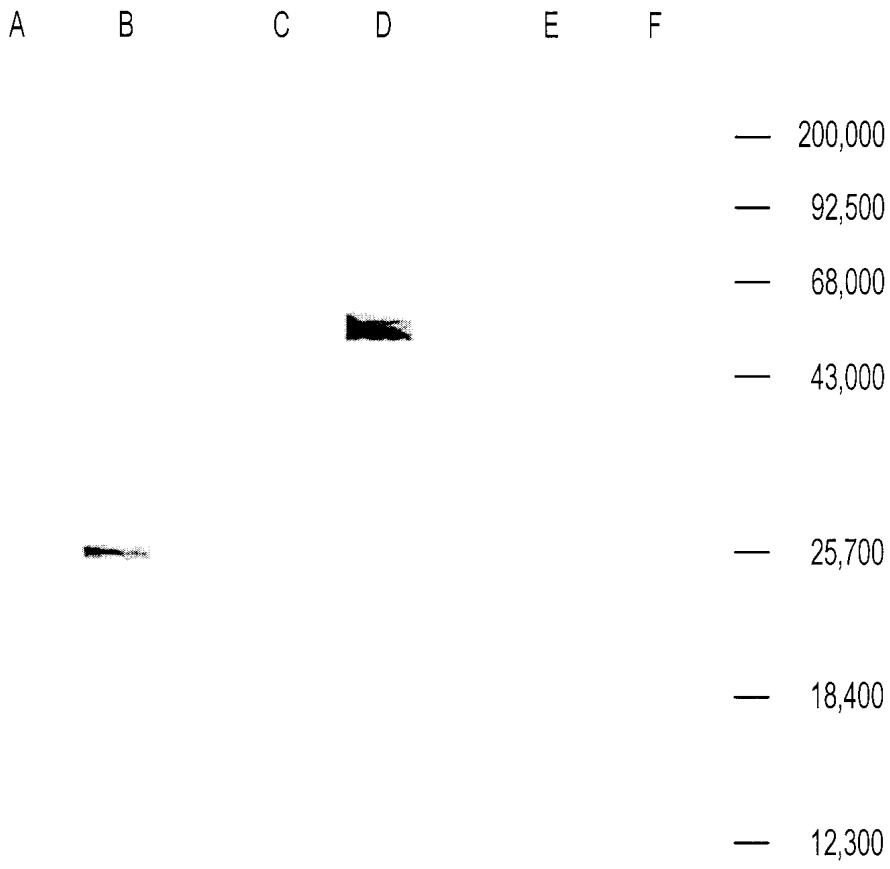


FIG. 30

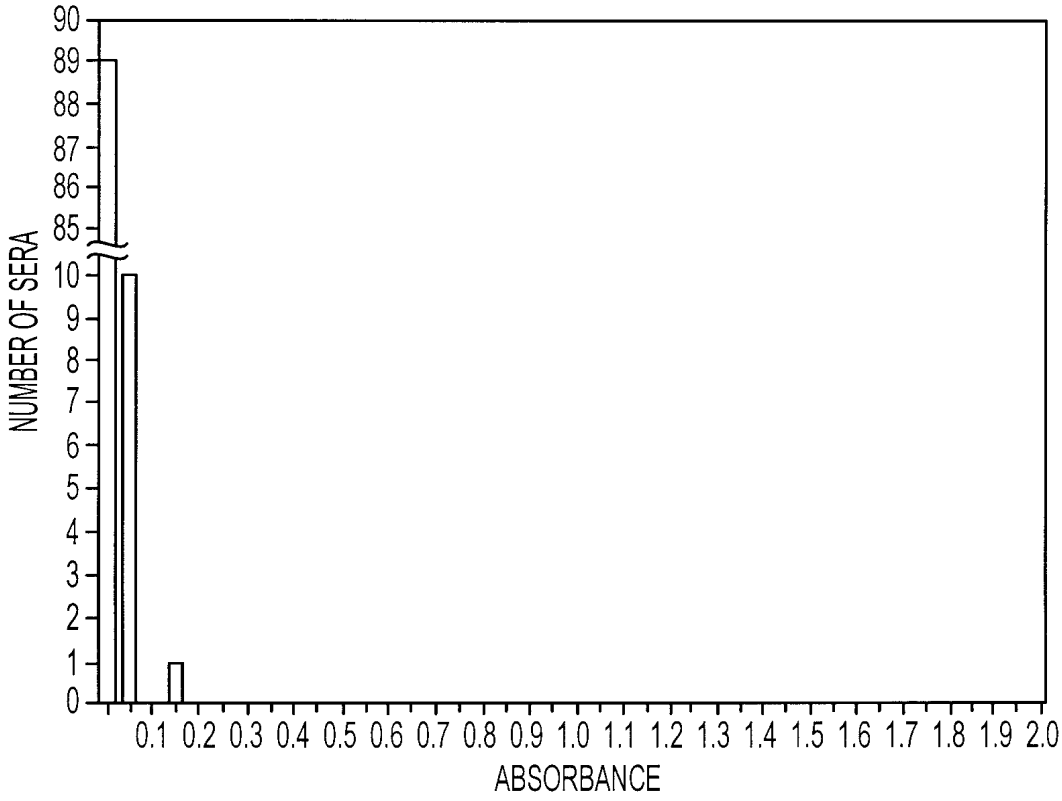


FIG. 31A

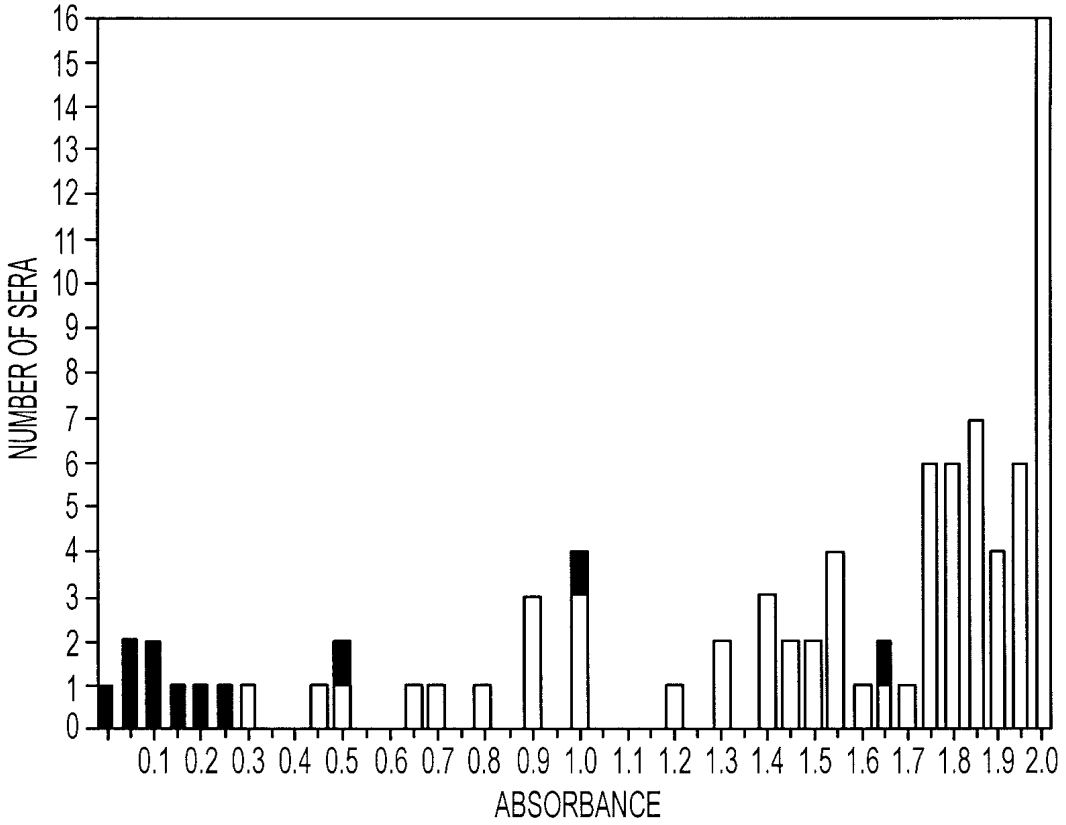


FIG. 31B